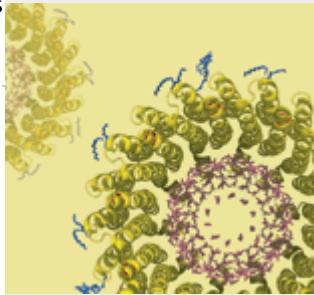


Table of Contents

29 April 2005
Volume 308
Number 5722



NEW THIS WEEK:

Sodium-Transporting ATP Rotors

Designer Surface Plasmons Confirmed

Building Successful Teams

Health, Productivity, and GM Rice

Understanding Intentions

▶ **Editors' Choice**

▶ **NetWatch**

▶ **ScienceScope**

▶ **Random Samples**

▶ **New Products**

▶ **Science Online Contents**

Special Feature

Research

This Week in Science

Reviews

Brevia

Research Articles

Reports

Technical Comments

News

News Summaries

News of the Week

News Focus

Commentary

Editorial

Letters

Policy Forum

Book Reviews

Perspectives

SPECIAL FEATURE

EDITORIAL FEATURE:

Not Your Father's Postdoc

Beryl Lieff Benderly 717-718.

RESEARCH

This Week in *Science*

Linked Rings from a Library * Understanding Others' Intentions * Nudging Optical Beams * A Salty Tropical Mix * Climate Clues from Glaciers * A Fly's Response to Climate Change * Societal Pressures and Primate Health * Designer Surface Plasmons * Going Through the Ring * Team-Building Exercise * Genetically Modified Rice in the Field * Both Waves * Keeping Time * Functional Brain Imagery 597

Editors' Choice: Highlights of the recent literature

CELL BIOLOGY: Sideline Quality Control * IMMUNOLOGY: Dendritic Cells, Part 1 * GEOCHEMISTRY: Preserved in Salt * OCEAN SCIENCE: A Shipping Forecast * SURFACE SCIENCE: Not-So-Thermal Desorption * MICROBIOLOGY: Dendritic Cells, Part 2 * PSYCHOLOGY: Traits in Common 603

Review

The Influence of Social Hierarchy on Primate Health

Robert M. Sapolsky 648-652.

Brevia

Horsfield's Hawk-Cuckoo Nestlings Simulate Multiple Gapes for Begging

Keita D. Tanaka and Keisuke Ueda 653.

Research Article

Structure of the Rotor of the V-Type Na⁺-ATPase from *Enterococcus hirae*

Takeshi Murata, Ichiro Yamato, Yoshimi Kakinuma, Andrew G. W. Leslie, and John E. Walker 654-659.

Structure of the Rotor Ring of F-Type Na⁺-ATPase from *Ilyobacter tartaricus*

Thomas Meier, Patrick Polzer, Kay Diederichs, Wolfram Welte, and Peter Dimroth 659-662.

Parietal Lobe: From Action Organization to Intention Understanding

Leonardo Fogassi, Pier Francesco Ferrari, Benno Gesierich, Stefano Rozzi, Fabian Chersi, and Giacomo Rizzolatti 662-667.

Reports

Amplification of Acetylcholine-Binding Catenanes from Dynamic Combinatorial Libraries

Ruby T. S. Lam, Ana Belenguier, Sarah L. Roberts, Christoph Naumann, Thibaut Jarrosson, Sijbren Otto, and Jeremy K. M. Sanders 667-669.

Experimental Verification of Designer Surface Plasmons

Alastair P. Hibbins, Benjamin R. Evans, and J. Roy Sambles 670-672.

All-Optical Switching in Rubidium Vapor

Andrew M. C. Dawes, Lucas Illing, Susan M. Clark, and Daniel J. Gauthier 672-674.

Extracting a Climate Signal from 169 Glacier Records

J. Oerlemans 675-677.

Early Local Last Glacial Maximum in the Tropical Andes

Jacqueline A. Smith, Geoffrey O. Seltzer, Daniel L. Farber, Donald T. Rodbell, and Robert C. Finkel 678-681.

Laboratory Earthquakes Along Inhomogeneous Faults: Directionality and Supershear

Kaiwen Xia, Ares J. Rosakis, Hiroo Kanamori, and James R. Rice 681-684.

Enhanced Diapycnal Mixing by Salt Fingers in the Thermocline of the Tropical Atlantic

R. W. Schmitt, J. R. Ledwell, E. T. Montgomery, K. L. Polzin, and J. M. Toole 685-688.

Insect-Resistant GM Rice in Farmers' Fields: Assessing Productivity and Health Effects in China

Jikun Huang, Rulifa Hu, Scott Rozelle, and Carl Pray 688-690.

A Rapid Shift in a Classic Clinal Pattern in *Drosophila* Reflecting Climate Change

P. A. Umina, A. R. Weeks, M. R. Kearney, S. W. McKechnie, and A. A. Hoffmann 691-693.

PERIOD1-Associated Proteins Modulate the Negative Limb of the Mammalian Circadian Oscillator

Steven A. Brown, Juergen Ripperger, Sebastian Kadener, Fabienne Fleury-Olela, Francis Vilbois, Michael Rosbash, and Ueli Schibler 693-696.

Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance

Roger Guimerà, Brian Uzzi, Jarrett Spiro, and Luís A. Nunes Amaral 697-702.

The Dynamics of Interhemispheric Compensatory Processes in Mental Imagery

A. T. Sack, J. A. Camprodon, A. Pascual-Leone, and R. Goebel 702-704.

Technical Comments**Comment on "Ecosystem Properties and Forest Decline in Contrasting Long-Term Chronosequences"**

Kanehiro Kitayama 633.

Response to Comment on "Ecosystem Properties and Forest Decline in Contrasting Long-Term Chronosequences"

D. A. Wardle, L. R. Walker, and R. D. Bardgett 633.

COMMENTARY**Editorial****Benefits of Women in Science**

Julia King 601.

Letters

A Cry for Help from Kansas *Eric Reynolds*; What Causes Lesions in Sperm Whale Bones? *Bruce M. Rothschild*; *Edward D. Mitchell*; *Michael J. Moore*, and *Greg A. Early*; Ethics of Tobacco Company Funding *Jed E. Rose*; Merits of a New Drug Trial for ALS? *Ettore Beghi*, *Caterina Bendotti*, *Tiziana Mennini*; *Tim Miller*, and *Don Cleveland*; CORRECTIONS AND CLARIFICATIONS 631.

Policy Forum**ECOLOGY:****Synthesizing U.S. River Restoration Efforts**

E. S. Bernhardt, M. A. Palmer, J. D. Allan, G. Alexander, K. Barnas, S. Brooks, J. Carr, S. Clayton, C. Dahm, J. Follstad-Shah, D. Galat, S. Gloss, P. Goodwin, D. Hart, B. Hassett, R. Jenkinson, S. Katz, G. M. Kondolf, P. S. Lake, R. Lave, J. L. Meyer, T. K. O'Donnell, L. Pagano, B. Powell, and E. Sudduth 636-637.

Books et al.**HIGHER EDUCATION:****Ideals Drowned in the Marketplace**

Harold T. Shapiro 634-635.

MICROBIOLOGY:**Appreciating Our Usual Guests**

Elaine Tuomanen 635.

Books Received 635.

Perspectives**APPLIED PHYSICS:****Toward Bridging the Terahertz Gap with Silicon-Based Lasers**

Alexander Borak 638-639.

SOCIOLOGY:**Network Theory--the Emergence of the Creative Enterprise**

Albert-László Barabási 639-641.

OCEAN SCIENCE:**Enhanced: Ocean Mixing in 10 Steps**

Bill Merryfield 641-642.

STRUCTURAL BIOLOGY:

Nature's Rotary Electromotors

Wolfgang Junge and Nathan Nelson 642-644.

NEUROSCIENCE:

Understanding Intentions: Through the Looking Glass

Kiyoshi Nakahara and Yasushi Miyashita 644-645.

NEWS

News of the Week

NASA:

Life Science Research on Space Station Is Headed for Big Cuts

Andrew Lawler 610-611.

NATIONAL ACADEMIES:

Panel Would Entrust Stem Cell Research to Local Oversight

Constance Holden and Gretchen Vogel 611.

NUCLEAR FUSION:

Tabletop Accelerator Breaks 'Cold Fusion' Jinx But Won't Yield Energy, Physicists Say

Charles Seife 613.

EARTH SCIENCES:

Earth Observation Program 'At Risk,' Academy Warns

Andrew Lawler 614-615.

DEPARTMENT OF ENERGY:

Falling Budget Could Force Choice Between Nuclear Science Facilities

Charles Seife 615.

SCIENCE POLICY:

Marburger Asks Social Scientists for a Helping Hand in Interpreting Data

Jeffrey Mervis 617.

U.S. PUBLIC SECTOR:

Agency Kills New Performance Rules

Jeffrey Mervis 617.

News Focus

PUBLIC HEALTH:

High Hopes and Dilemmas for a Cervical Cancer Vaccine

Jon Cohen 618-621.

PUBLIC HEALTH:

HPV's Peculiarities, From Infection to Disease

Jon Cohen 619.

INFORMATION SHARING:

Europe Steps Into the Open With Plans for Electronic Archives

Gretchen Vogel and Martin Enserink 623-624.

DEVELOPMENTAL BIOLOGY:

Combing Over the Polycomb Group Proteins

Jean Marx 624-626.

U.S. EDUCATION:

Kansas Gears Up for Another Battle Over Teaching Evolution

Yudhijit Bhattacharjee 627.

Departments

AAAS News & Notes 646-647.

Products

NEW PRODUCTS 705.

NetWatch

NET NEWS: Species Master List Hits Milestone * WEB TEXT: Living Small * RESOURCES: Precautionary Principles *

EDUCATION: Math Motherlode * IMAGES: Retracing a Long Walk 609

ScienceScope

Congress Probes Charges of Harassment at NIH * Two Israeli Universities Targeted for Boycott * Griffin Moves Fast To Reshape NASA * Netherlands Reports First vCJD Case * UC Retains Oversight of Lawrence Berkeley 613

Random Samples

Monkeys Strike Gold * New Cambridge Center Emerges * Sex and Science (cont.) * Champion Racer Cloned * Awards * In The News * Pioneers * Snafus 628

Understanding Others' Intentions

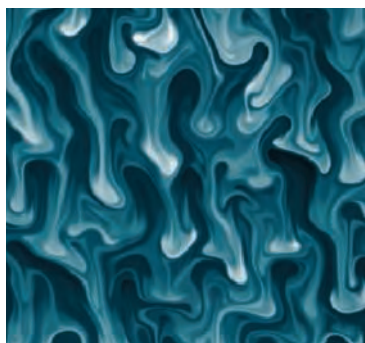
When we act, we intend to reach a goal. Conversely, when we observe someone else act, we can often infer their intentions. **Fogassi et al.** (p. 662; see the Perspective by **Nakahara and Miyashita**) found that in the inferior parietal lobule of an individual about to begin an action, the goal of their action (e.g., grasping for food versus grasping a branch) is reflected in the discharge of the neurons coding the first element of the sequence leading to the goal. In addition, many parietal neurons that code for an action like grasping also discharge while watching someone else grasping (parietal mirror neurons). The majority of these neurons respond differentially when the same observed motor act is performed with a different goal. Thus, these mirror neurons, besides describing the observed motor activity, also predict the intention behind the action.

Nudging Optical Beams

Most optical switching takes place with mirrors or electro-optic devices. Some applications, however, might be better served with all-optical technology, where light in one beam controls another. **Dawes et al.** (p. 672) report the use of rubidium vapor as an optical switching medium. Strong laser beams interacting in the vapor create multiple exit beams, and these beams can be rotated by applying a much weaker control beam.

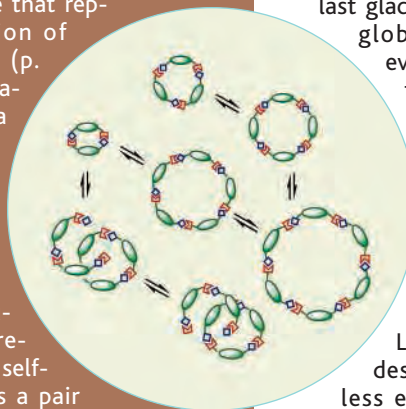
A Salty Tropical Mix

Diapycnal mixing, which occurs between adjacent layers that stratified because of density differences, can control the distribution of heat, carbon dioxide content, and numerous other properties of the ocean. Double-diffusion, such as by the formation of salt fingers, is a mechanism by which this type of mixing can be enhanced, but which is unquantified over most of the ocean. **Schmitt et al.** (p. 685; see the Perspective by **Merryfield**) present results from a large-scale ocean tracer experiment that covered 1.3 million square kilometers of the tropical Atlantic Ocean.



Linked Rings from a Library

Combinatorial chemical synthesis can rapidly generate many different compounds, but they often share an underlying structure that represents a fairly small region of chemical space. **Lam et al.** (p. 667) used a dynamic combinatorial approach to discover a surprisingly elaborate structure for binding the acetylcholine neurotransmitter. The authors added synthetic dipeptides to an acetylcholine solution under conditions allowing reversible coupling. At equilibrium, the predominant structure that had self-assembled as a receptor was a pair of linked 42-membered rings, each a trimer of the dipeptide building blocks. This catenane molecule was isolated in 65% yield and showed a 100-nanomolar affinity for the neurotransmitter.



Climate Clues from Glaciers

Direct instrumental records have shown that average surface temperatures have risen significantly across the globe during the past two centuries. Glaciers have responded to this warming, mostly by retreating, and changes in the extents of glaciers typically have been understood and modeled as a function of the temperature of the overlying atmosphere. **Oerlemans** (p. 675, published online 3 March 2005) has reversed this order. By analyzing a large set of data on glacier length fluctuations dating back to the mid-17th century, he has reconstructed an independent record of temperature variability and found that global warming began earlier (in the middle of the 19th century) than in other temperature reconstructions. Was the last glacial maximum (LGM) a globally synchronous event, or did it ripple in time across the world in a more complex way? **Smith et al.** (p. 678) present a suite of cosmogenic ^{10}Be ages from moraines in Peru and Bolivia which show that the local LGM in the tropical Andes occurred earlier and less extensive than previously believed. Glaciers reached their terminal position about 34,000 years before present, long before the date of 21,000 years before present often assigned to the LGM, and terminated at positions much higher up-valley than did larger previous glaciers. Their findings imply that the decrease in tropical temperatures there was only half that of most other estimates of 6° to 7°C.

A Fly's Response to Climate Change

Clinal variations are in genetic polymorphisms that occur across an organism's geographical range as allele frequencies change with climate gradients. A classic example is the cline in the alcohol dehydrogenase (*Adh*) gene of the fly *Drosophila melanogaster* from north to south on the east coast of Australia. **Umina et al.** (p. 691) characterized this cline in a large number of populations ranging from the wet tropics to cool temperate regions along the entire coast. An abrupt shift was observed in the elevation of the *Adh*^S allele during the past 20 years, when marked change occurred in several climatic variables along the cline. The drier and warmer climate of recent years is likely to account for the change in the cline, emphasizing how the genetic composition of populations could change in response to climate even in widespread species that are adapted to a range of climatic conditions.

CONTINUED ON PAGE 599

Societal Pressures and Primate Health

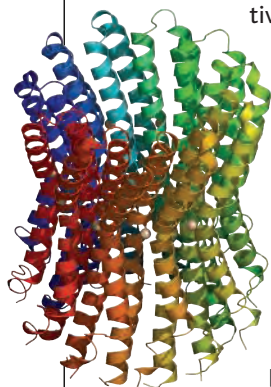
Primate populations, including humans, are organized in various ways, but usually include dominance hierarchies. **Sapolsky** (p. 648) reviews how, depending on the specific features of each society, it may be the lower ranking or the higher ranking members of the society that experience greater stress. This dominance-related stress produces physiological changes that ultimately are detrimental to the individual's health. The principles that emerge from studies on nonhuman primates about dominance effects on health may also apply to humans.

Designer Surface Plasmons

Conducting metal films usually are not expected to support surface plasmon modes, which are localized excitations of electrons coupled with electromagnetic radiation. However, recent theoretical work has predicted that these bound electromagnetic states could be induced on conducting surfaces by perforating them with holes. Working in the microwave regime, **Hibbins et al.** (p. 670) verify that surface plasmon-like modes can indeed be induced by controlling the geometry of the metallic sample. The ability to tune or design these surface modes may have consequences for applications involving the propagation of surface plasmons.

Going Through the Ring

The ion-transporting adenosine triphosphates (ATPases) of the F-type (e.g., mitochondrial proton ATPase) and of the V-type (e.g., vacuolar proton ATPases) have roughly similar overall structures, with a threefold symmetric F_1 (or V_1) portion and an integral membrane ring (F_0 or V_0) of anywhere from 10 to 14 identical subunits. Some of these enzymes transport Na^+ instead of protons (see the Perspective by **Junge and Nelson**). **Murata et al.** (p. 654, published online 31 March 2005; see the cover) present a high-resolution structure of the 10-subunit ring of a Na^+ -transporting V-type ATPase. Each subunit contributes four transmembrane helices to a ring of about 83 angstroms in diameter, and the Na^+ binding site is exposed on the outer surface of the ring, about midway into the membrane bilayer. **Meier et al.** (p. 659) present a high-resolution structure of the 11-subunit ring of a Na^+ -transporting F-type ATPase, in which each subunit contributes only two transmembrane helices to a smaller ring of about 50 angstroms in diameter. Both structures are consistent with a model in which ATP-driven rotation of the ring causes a bound Na^+ to be ejected to the outside, which is then followed by refilling of the transport site by a Na^+ from the inside.



Team-Building Exercise

What are the factors required to build a successful creative team? **Guimerà et al.** (p. 697; see the Perspective by **Barabási**) used network analyses to model such factors and found a clear relation between team diversity, collaboration network structure, and team performance. Within a scientific discipline, greater journal impact factor correlates strongly with larger teams, a lower tendency to "over-repeat" collaborations, and significant presence of both experienced researchers and newcomers. Similar properties appear to have contributed during the last century to define the most successful team composition for Broadway musical productions.

Genetically Modified Rice in the Field

China has developed rice strains that are genetically modified to be intrinsically resistant to pests, and **Huang et al.** (p. 688) describe preliminary field trials carried out in 2002 and 2003 with these strains. For plots planted with pest-resistant genetically modified rice strains, the farmers could reduce their use of pesticides by as much as 80%. At the same time, yields increased, and the health of the farmers improved significantly with reduced occupational exposure to pesticides.

CREDIT: MEIER ET AL.

ADVERTISER DIRECTORY

advances in:

Proteomics

Analysis and Automation

The following organizations have placed ads in the Special Advertising Section

Advances in:

Proteomics

Analysis and Automation

ADVERTISER	Page
Elchrom Scientific AG	711 & 713
Fuji Photo Film Co., Ltd.	709
Leica Microsystems GmbH	706
Syngene	714

Turn to page 707

Benefits of Women in Science

Recent comments from Harvard President Lawrence Summers have sparked heated discussion in the United States and abroad about possible inherent (that is, genetic) differences between women and men. The debate concerns whether these differences might explain the paucity of women in elite science, engineering, and technology (SET) careers. The issue really amounts to possible differences at the high extremes of ability distributions, but the available evidence is that any inherent differences are swamped by social and cultural factors. It is the failure to encourage more women to pursue SET careers, and to maintain their presence in these positions, that requires serious attention. As John Brock, the chief operating officer of Cadbury Schweppes, points out “A diverse workforce . . . is the best way to expand into new markets and stimulate new business ideas . . . that’s a significant competitive advantage.”

In the United Kingdom, we have a pressing need to encourage more women to enter SET careers. The UK government’s agenda for economic growth includes a commitment to increase the proportion of gross domestic product spent by both government and industry on R&D. Yet the Institute of Employment Studies predicts that by 2011, only 20% of the workforce will be white, male, able-bodied, and under 45. Eighty percent of future employment growth will be attributable to women.

Industry has recognized the value of an experienced female staff. In 2002, Lord Browne, chief executive of British Petroleum (BP), remarked that “because the management of the industry has been predominantly white and male and Anglo-Saxon, those people have recruited and promoted in their own image.” Among other initiatives, BP has appointed a Vice President for Diversity, and Shell Oil holds recruiting events for female engineers at UK universities. Support for female employees during career breaks is becoming more common in UK-based companies, as industry recognizes that diversity is a strategic business issue. Industry has also responded to research showing that diverse teams are harder to manage than homogeneous groups: Absenteeism and staff turnover are higher; communication and social integration take more effort; common values and rules must be established; and the different needs, behaviors, and characteristics of team members must be supported. Team leaders must learn to manage differences of opinion—the very source of the diversity advantage. But the results are worth having: Diverse teams outperform on innovation, problem-solving, flexibility, and decision-making.

The UK’s Athena program was established in 1999 to address the shortage of women in SET academic careers and to deliver a significant increase in the number of women recruited to top academic jobs. The Athena Survey of SET (ASSET) report (just released) compares career pathways of more than 6500 men and women in academia and research institutes in the United Kingdom.* The survey reveals that differences between women’s and men’s experiences are more marked in academia than in other kinds of research organizations. Men in academic positions are more likely to report that they were encouraged to apply for promotion, as compared with their female colleagues. In academia, women rank annual performance reviews and personal development more highly than men in supporting career progression; in research institutes, the ranking by both sexes is almost identical. Nearly 50% of women in universities feel disadvantaged in terms of salary and promotion, whereas only 15% of male staff recognize this as a problem for their female colleagues.

This is not to say that things haven’t improved. When I went up to Cambridge University in the 1970s as an undergraduate, only 16% of all undergraduates were female, with a mere 2% studying physical sciences, and there were no female academic staff in the departments of physics, chemistry, materials science, engineering, or mathematics. Now, Cambridge University has about 49% women undergraduates, of which 10 to 25% study the physical sciences, and 24% of the academic staff in the materials science department are women. At Imperial College (London), our fastest growing engineering course is bioengineering, with an undergraduate intake of 50% women.

Academic research and initiatives such as Athena have been effective in highlighting the benefits of diversity and the management challenges of maintaining a diverse workforce. Industry sees the competitive and financial advantages and has responded. Despite showing the way, academia is being left behind. We must embed in our universities the best practices that we preach.

Julia King

Julia King is Principal of the Engineering Faculty at Imperial College, London.

*See www.athenaproject.org.uk.



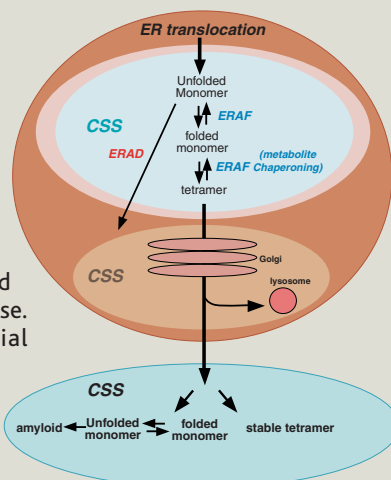
edited by Gilbert Chin

CELL BIOLOGY

Sideline Quality Control

Quality control within the endoplasmic reticulum has long been regarded as a mechanism that prevents the secretion of misfolded proteins: Endoplasmic reticulum-associated degradation (ERAD) and the inability of incorrectly folded proteins to access the export machinery are its key factors. However, in some cases, quality-control mechanisms fail, and misfolded or misassembled proteins are secreted and cause disease. One class of such diseases is known as the familial amyloidoses, in which aberrant forms of the protein transthyretin are secreted, become misfolded, and form pathological aggregates.

Sekijima *et al.* have examined the thermodynamics and kinetics of the folding and assembly of disease-associated forms of transthyretin. The endoplasmic reticulum is the entry site of the protein secretory pathway, and export from this compartment allows aberrant or misfolded proteins to transit to the Golgi and beyond. For many mutant forms of transthyretin, the balance between endoplasmic reticulum-assisted folding (ERAF) and ERAD determines the overall performance of this gatekeeping stage, and some cell types can actually secrete aberrant transthyretin efficiently. The competition between these intracompartamental pathways defines the ability of a particular type of cell or tissue to restrict or permit the secretion of aberrant proteins, and thereby determines the tissue selectivity and severity of protein-folding disorders. — SMH



Model for how the competitive stability score (CSS) predicts the partitioning between ERAF and ERAD.

Cell 121, 73 (2005).

that the inclusion in the salt crystal, and hence the bacterium, might be a contaminant of an uncertain and possibly younger age; the retention of younger fluids flowing through or adjacent to older rock is not uncommon.

Satterfield *et al.* have now determined the chemistry of the fluid inclusions in these salt crystals. Earth's ocean chemistry has changed over time, and the Late Permian oceans were depleted in Mg and sulfate as compared with today's oceans, which provides a signature that is diagnostic for this time period. The chemistry of the inclusions fits with that of Permian seawater, suggesting that the bacterium is indeed old. — BH

Geology 33, 265 (2005).

OCEAN SCIENCE

A Shipping Forecast

Phytoplankton are responsible for 45% of plant primary production and absorb large amounts of carbon dioxide from the atmosphere. From 1997 to 2002, the satellite-based Sea-viewing Wide Field-of-view Sensor (SeaWiFS) collected global data on the distribution of chlorophyll *a*, a measure of phytoplankton concentration; however, this data set is too short to provide insights into decadal changes in phytoplankton.

Raitsos *et al.* have turned therefore to measurements of the phytoplankton color index, which have been collected since 1931 along shipping routes in the North Sea and the North Atlantic and which have used a consistent sampling and measurement methodology since 1948. The authors demonstrate a significant correlation between the two data sets from 1997 to 2002 and then use this correlation to retrospectively calculate monthly changes in chlorophyll *a*

IMMUNOLOGY

Dendritic Cells, Part 1

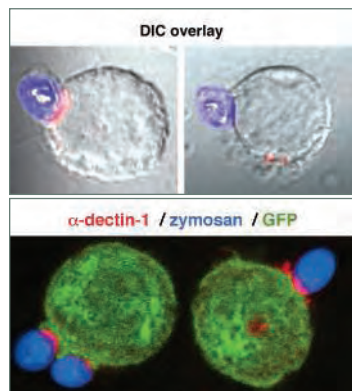
The recognition of the molecular patterns of pathogens by innate immune receptors is a well-established function of the Toll-like receptor (TLR) family; similar activities are now being ascribed to other families of host cell proteins. For example, the C-type lectin Dectin-1 enables phagocytosis of yeast by scavenger cells by binding the yeast cell wall carbohydrates (β glucans), and it has been shown to act as a coreceptor for TLR2, leading to inflammatory cytokine expression.

Rogers *et al.* show that Dectin-1 can signal directly to initiate cytokine transcription. The production of interleukin-2 (IL-2) and IL-10 could be induced upon exposure of dendritic cells to a yeast cell wall extract

and was partially blocked by a soluble β glucan. An equivalent phenotype could be conferred on a B cell hybridoma line (LK cells) by transduction of Dectin-1. Transcription of both cytokines was dependent on the intracellular tyrosine kinase, Syk, which was recruited by the immunoreceptor tyrosine-based activation motif in the

cytoplasmic tail of Dectin-1. The distinct cytokine profiles induced by Dectin-1 in the context of Syk signals, versus co-signaling with TLR2, suggest flexibility in innate pattern recognition that could be tailored for a pathogen-specific adaptive immune response. — SJS

Immunity 22, 509 (2005).



Localization of yeast cell wall (blue) and Dectin-1 (below, red) and Syk (above, red) on the surface of LK cells.

GEOCHEMISTRY

Preserved in Salt

The most ancient living organism is claimed to be a bacterium that has been extracted and cultured from a small bubble of fluid trapped in a Permian-aged (~250 million years ago) salt crystal, similar to the way that, for example, insects are trapped in amber. The idea is that this bacterium became entombed in a fluid inclusion in the salt crystal and remained dormant until it was resuscitated. One criticism has been

CONTINUED ON PAGE 605

concentrations since 1948. The results show a marked increase in chlorophyll *a* in the mid-1980s, a time when the composition and productivity of the regional ecosystem are known to have changed. This data set will be useful for biogeochemical and climate modeling studies that aim to understand the links between marine biology and climate. — JFU

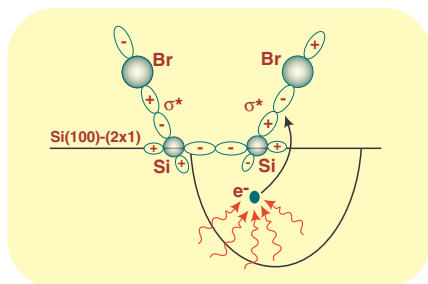
Geophys. Res. Lett. **32**, 10.1029/2005GL022484 (2005).

SURFACE SCIENCE

Not-So-Thermal Desorption

The desorption of atoms or molecules from surfaces is thought to proceed through one of two mechanisms. Heating of the surface usually results in thermal desorption, in which the bonds holding the adsorbed species are put into such high vibrational states that they break. In electron- or photon-stimulated desorption, excitation of the adsorbate into an antibonding electronic state leads to desorption.

Trenhaile *et al.* followed the desorption of Br from the Si(100)-(2×1) surface at 620 to 775 K via scanning tunneling microscopy. Their analysis shows that this process does not proceed through vibrational excitation but by electron capture into long-lived states that then populate an



A schematic model of the desorption process.

antibonding σ^* state that then ejects the Br atom. The excitation energy for desorption changes with the Fermi level for different silicon doping levels. Entropy can actually help drive this process, in which 10 to 20 optical phonons come together to push the electron over its barrier. — PDS

Surf. Sci. **10.1016/j.susc.2005.3.053** (2005).

MICROBIOLOGY

Dendritic Cells, Part 2

The first step to infection is capture by a cell-surface receptor. A broad range of viruses, bacteria, and other human pathogens initiate infection by attaching to dendritic cell-specific ICAM-3 grabbing

nonintegrin (DC-SIGN), a C-type lectin encoded by the gene *CD209*. The usual role of DC-SIGN is to mediate contact between dendritic cells and T cells and to promote the migration of dendritic cells through tissues.

Sakuntabhai *et al.* have explored the effects of genetic variation in *CD209* on the specific disease syndromes caused by dengue virus. They recruited school-aged children with classical incapacitating dengue fever from three hospitals in Thailand. Screening *CD209* for genetic polymorphisms revealed that a dominant protective effect against dengue fever (without leakage of plasma), but not dengue hemorrhagic fever, lay in a G allele in the promoter region of *CD209*. This polymorphism influences the binding of the transcription factor Sp1 and may ultimately affect disease progression as well as the distinct pathophysiologies of dengue fever and dengue hemorrhagic fever. — CA

Nat. Genet. **10.1038/ng1550** (2005).

PSYCHOLOGY

Traits in Common

The five-factor model of personality posits five basic dimensions of personality: neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. Previous cross-cultural work has relied primarily on self-report measures collected mostly from Westernized college students. Using a third-person form of the NEO personality inventory, McCrae *et al.* present an intercontinental analysis of observer ratings. College students were asked to rate individuals from one of four groups—college-aged men and women and adult men and women—on six facets in each of the five dimensions. They find that their model does appear to apply across all 50 cultures (including Arabic and black African); the fit isn't perfect, but some of the variation may be due to mismatches between the questionnaire items and cultural contexts. Women were generally rated more highly than men, confirming data from self-report inventories, and scored higher on all six facets of agreeableness, which is consistent with earlier observations that women are more lenient when rating others. One interesting trend is that adult men scored higher than women on the conscientious facet "achievement striving," whereas the opposite ranking applied for college-age individuals, possibly reflecting a role reversal across generations. — GJC

J. Pers. Soc. Psychol. **88**, 547 (2005).

GetInfo

science.labvelocity.com



Get the lab
product info
you need
— FAST



Science announces a new online life science product information system, **GetInfo**, powered by **LabVelocity**

- Quickly find and request free information on products and/or services found in the pages of *Science Magazine*
- Ask vendors to contact you with information
- View detailed product information
- Link directly to vendors' websites

Visit GetInfo today at
science.labvelocity.com



NET NEWS

Species Master List Hits Milestone

An international project to create a comprehensive listing of life on Earth is about one-third complete. Last week, the latest update to the Catalogue of Life pushed the total number of species in this taxonomic trove to more than 535,000.

The catalog is sponsored by the Integrated Taxonomic Information System (ITIS) and Species 2000, a consortium of database organizations based at the University of Reading, U.K. (*Science*, 14 July 2000, p. 227). The Species 2000 site serves as a portal to the catalog, allowing you to browse or search a taxonomic tree linked to a host of "federated" databases such as AlgaeBase, the Species Fungorum, the World Spider Catalog, and many more. For example, look up the gerenuk (above), an African antelope, to find information such as its accepted scientific and common names, distribution, and classification. You can link to the ITIS database for more details. Smithsonian Institution zoologist Michael Ruggiero, director of ITIS, says the project is on track to record all of the roughly 1.75 million named species by 2011.

www.sp2000.org



RESOURCES

Precautionary Principles

Looking for data on the health risks of beryllium or advice about cleaning up spills of phthalic acid? Immerse yourself in chemical safety information at this site from the United Nations and other international organizations. The collection of fact sheets, reports, and other documents profiles hundreds of widely used substances and products, such as the flavoring zingerone, which gives gingersnaps their snap. For a quick rundown on a chemical's risks, flip through the chemical safety cards. Longer documents evaluate hazards from specific pesticides, potential carcinogens, and other kinds of compounds.

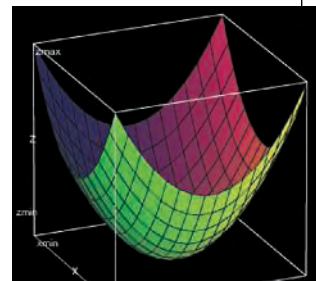
www.inchem.org

EDUCATION

Math Motherlode

Math teachers looking for a telling example or lucid graphics to jazz up their classes can check out this Web site from the Mathematical Association of America. The online library furnishes tools, animations, and other resources to help high schoolers and undergraduates hone their math skills. Exercises let users do everything from graphing 3D equations to investigating the scatter of German rocket strikes on London during World War II, a classic example of the pattern called the Poisson distribution. With open-source math applets called Osslets, students can sink their teeth into topics such as linear transformation. The site also houses a journal with articles on using history to teach math—for example, analyzing paintings by Leonardo da Vinci and other Renaissance artists can help students understand geometry.

www.mathdl.org/jsp/index.jsp



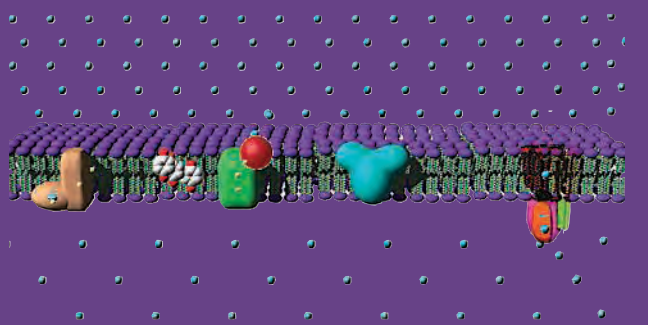
WEB TEXT

Living Small

A restless throng of hydrogen ions lurks above a bacterial membrane. Pumped out by the cell, the ions push back across the membrane and turn molecular turbines (rightmost structure) that fashion ATP to

power the microbe. Students can discover more about how a bacterium works at this online microbiology textbook from Tim Paustian of the University of Wisconsin, Madison. Still under construction, the site includes 17 partial or complete chapters covering everything from bacterial structure and nutrition to viral pathogens like the pesky rhinoviruses that cause colds. The text weaves in plenty of animations and fun tidbits, such as a section on the hardy *Pseudomonas* bacteria that can eat nitroglycerin and TNT. Paustian also comments on bugs in the news, including the bird flu outbreak in Southeast Asia.

www.bact.wisc.edu/Microtextbook



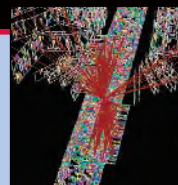
IMAGES

Retracing a Long Walk

Earlier this month, the National Geographic Society and IBM announced a project to produce a sharper picture of human migrations by analyzing DNA samples from 100,000 people (*Science*, 15 April, p. 340). The Web site of the Genographic Project is worth a look for the lavishly illustrated backgrounders on genetics and migrations. A timeline depicts what we know about the human expansion from Africa beginning about 60,000 years ago, stopping at landmarks such as the controversial Cactus Hill site in Virginia. Evidence found there suggests that people reached the Americas thousands of years earlier than previously thought. Another section explains how to send in your DNA and find out where your ancestors originated. Genealogical curiosity will cost you \$99.95 plus shipping for the test kit.

www5.nationalgeographic.com/genographic

Send site suggestions to netwatch@aaas.org. Archive: www.sciencemag.org/netwatch



NASA

Life Science Research on Space Station Is Headed for Big Cuts

NASA is putting the finishing touches on a new plan to slash the quality and quantity of cutting-edge research on the international space station. The space agency intends to postpone and cancel a number of experiments, abandon a host of research facilities, and reduce the amount of crew time and agency funding devoted to station science, according to outside scientists and NASA officials familiar with the plan. Scientists are also upset that they have been largely excluded from the review, and politicians are complaining about the apparently shrinking payoff from the billions being spent on the orbiting laboratory.

The revamped research plan follows President George W. Bush's call last year for NASA to step up work on lunar and Mars exploration. That redirection of the space program would dedicate the station to collecting life sciences data that would benefit astronauts living and working for long periods beyond Earth orbit. But the cost of returning the shuttle to flight, combined with the rush to finish the station by 2010 and build new launchers, is forcing the agency to put the squeeze on what would appear to be priority research in biology, along with several science missions not connected to the station (see p. 614 and *Science*, 22 April, p. 484).

One major change would eliminate animal research facilities—including a centrifuge, regarded as the centerpiece of the life sciences effort, now under construction in Japan—and virtually end basic biological research. Instead, U.S. station research would consist primarily of experiments using astronauts as test subjects. NASA documents also show that the agency is planning to reduce the number of racks that hold experiments, the funding to prepare those racks for orbit, and the hours astronauts devote to research in space.

This limited science portfolio is a far cry from former President Ronald Reagan's 1984 description of astronauts achieving "quantum leaps" in science, communications, materials, and medicine. That retreat worries some U.S. lawmakers. "I want to go



Time flies. Astronauts may have fewer hours in which to do research aboard the station.

back to the Ronald Reagan vision," declared Senator Kay Bailey Hutchison (R-TX), chair of a panel with NASA oversight, during a hearing last week on station research. "This impressive facility cannot be allowed to be used simply as a tool for moon and Mars exploration-related research."

That concern is bipartisan and global. Another member of the committee, Senator Bill Nelson (D-FL), said that he and Hutchison "are of one mind" on the matter. Dieter Isakeit, a spokesperson for the European Space Agency (ESA), says his organization will stay the course with its research program, which covers many disciplines in the physical

and life sciences. Japanese officials, meanwhile, say that they expect to discuss the station design and research program during a fall meeting with the space station partners.

Notwithstanding those concerns, NASA appears unlikely to return to the original research vision for the station. Commercial interest in studies relating to drug discovery never gelled, for example, and in the late 1990s, NASA began tapping funds for research facilities to pay for station cost overruns. Work in the materials sciences was largely jettisoned after a 2002 review, and the 2003 Columbia disaster severely curtailed short-term research plans.

Meanwhile, NASA managers "are finding other things more pertinent" to fund than science, says Kenneth Baldwin, a biologist at the University of California, Irvine. And it's making those decisions largely on its own. "The science community is basically out of the loop," says Baldwin, who chaired the agency's biological and physical sciences advisory panel, which was abolished last year as part of a general advisory council reorganization. The science panel likely will become part of an exploration committee chaired by retired Air Force General Lester Lyles.

Baldwin says the space biology effort would be "decimated" in the new plan. Both he and Charles Oman, a Massachusetts Institute of Technology (MIT) aerospace engineer tracking the research plan, expect that the animal research facilities will be dropped. In addition, documents first posted last week by the Web site NASAWatch show that the agency will roughly halve the number of station racks in use aboard the space station to four; limit astronaut hours from the 15 hours planned to 10 hours; and slice funding for integrating the experiments into the racks by 38% starting in 2006. NASA Deputy Chief Scientist Howard Ross says that the document, to be completed next month, is only "for planning purposes." And he rejects the notion that the community has been excluded from discussions.

Meanwhile, station research scientists say they are waiting anxiously for word on what will fly. Physicist Sam Ting, a Nobel laureate at MIT, still hopes to launch his Alpha Magnetic Spectrometer to the station in 2008 to search for antimatter. He says NASA paid only 5% of the \$1.2 billion cost of the project, which includes participants from 16 countries. If it's dropped, Ting says, "then I don't see how NASA can say it wants

CREDIT: JPL/NASA



international cooperation.”

Even those experiments that seem directly tied to humans living in space are not safe. An experiment proposed by Baldwin and European colleagues more than a year ago to examine the molecular biology of muscles in microgravity passed peer review by an international team of scientists and won NASA approval last year. But 3 weeks ago, Baldwin

received word that the project was being placed on indefinite hold.

NASA already has pulled the plug on a project by MIT and the Sorbonne University in Paris to test human spatial orientation and motor behavior in space. Oman, the principal investigator, says NASA decided to cease funding the project, called Voila, after the hardware was completed. Oman says he

is sympathetic to the challenges facing NASA in trying to balance flight hardware and science, and he applauds the concrete goals set by the president. But he doesn't hide his disappointment. “The station is not going to be the world-class facility we foresaw,” he says. “That is the cold reality.”

—ANDREW LAWLER

With reporting by Daniel Clery and Dennis Normile.

NATIONAL ACADEMIES

Panel Would Entrust Stem Cell Research to Local Oversight

The National Research Council and the Institute of Medicine this week called for the creation of a new layer of oversight at institutions where research on human embryonic stem (ES) cells is conducted.

The recommendation is part of guidelines* developed by an academy panel “in the absence of federal regulations specifically designed” for this research. The committee, headed by Richard O. Hynes of the Massachusetts Institute of Technology, cited as precedent the Asilomar conference of 1975. At that meeting, scientists formulated their own guidelines for recombinant DNA research, helping ease public fears about the new field.

But unlike Asilomar, which established the Recombinant DNA Advisory Committee at the National Institutes of Health because of the patchwork of laws across the United States, the new panel leaves many of the tough questions to local committees. The panel calls on every institution that hosts human ES cell research to set up an Embryonic Stem Cell Research Oversight (ESCRO) committee containing experts well versed in the scientific, medical, legal, and ethical questions. It “should not [just] be a subcommittee” of the existing institutional review board, the panel warns. The recommendation makes sense, says Irving Weissman of Stanford University, who was not on the panel: “These issues transcend the usual expertise of institutional review boards.”

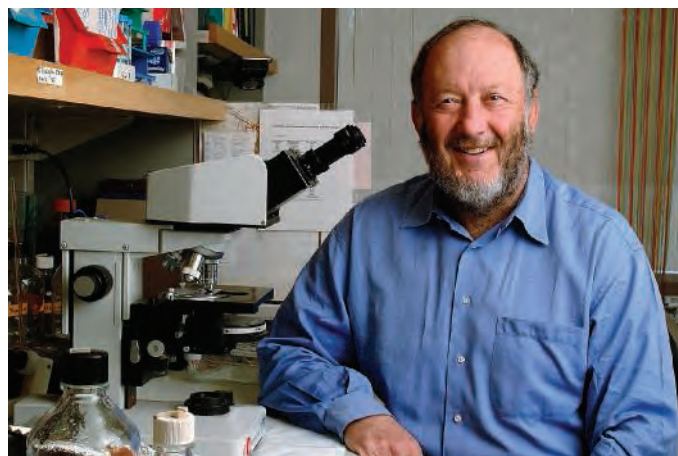
The main thrust of the 131-page report is procedural, not ethical. It rules out few kinds of research and leaves most decisions to the local committees. In addition to keeping track of all research involving human ES cells, the panels should review everything related to the derivation of new cell lines, whether created from leftover blastocysts from fertility treatments,

through nuclear transfer (otherwise known as research cloning), or “made specifically for research” by in vitro fertilization of donor sperm and egg. That last option “is controversial,” affirms stem cell researcher Evan Snyder of the Burnham Institute in La Jolla, California. Although many scientists agree on the desirability of nuclear transfer, they question the ethics of creating fertilized embryos “specifically” for research. “Nobody I know seriously entertains” that option, adds Snyder. Panel member Norman Fost, an ethicist at the University of Wisconsin, Madison, says the committee discussed

14 days old. The committee saw only limited potential in other approaches for generating cell lines that might bypass ethical difficulties (*Science*, 24 December 2004, p. 2174).

On the potentially controversial topic of using ES cells to create chimeras—animals that contain the genome of a different animal in some of their cells—the panel notes that “chimeras are widely used in research; ... thus there seem to be no new ethical or regulatory issues regarding chimeras themselves.” The panel points out that chimeras are valuable for testing the qualities of human ES cells. How-

ever, because pluripotent cells have the potential to turn into many kinds of cells, the committee says no animal ES cells should be injected into human blastocysts, and no human ES cells should be allowed into the blastocysts of other primates. And because ES cells can theoretically travel to the gonads and produce sperm and egg cells, no animal that has received human ES cells should be allowed to breed. That leaves



Go-ahead. An academy panel did not rule out Irving Weissman's proposed experiments that would inject human ES cells into mouse brains.

the issue but decided to leave the decision to local committees. “The requirement for new committees to oversee this kind of research ... reflects the seriousness of the issue,” he says.

The report dwells at length on the need for informed consent from donors of eggs, sperm, blastocysts, or somatic cells for ES cell research and says explicitly that donors should not be paid. It also confirms that no research should be allowed on embryos over

Weissman's “Stuart Little” mouse in the clear. Weissman has stirred controversy with his plan to grow brain cells from human ES cells in mice to study how the cells develop and make connections with each other.

The panel also recommends creation of a national body to periodically assess the adequacy of the guidelines and provide a forum for continuing discussion.

—CONSTANCE HOLDEN AND GRETCHEN VOGEL

CREDITS: STANFORD UNIVERSITY MEDICAL CENTER

* *Guidelines for Human Embryonic Stem Cell Research*, www.nap.edu/books/0309096537/html

Tabletop Accelerator Breaks 'Cold Fusion' Jinx But Won't Yield Energy, Physicists Say

A crystal with a strange property is at the heart of a clever method for inducing nuclear fusion in a tabletop-sized device. The inventors of the machine—which works by firing fine beams of atomic nuclei at other atoms—are not billing it as a possible source of energy, but they say it could serve as a portable source of neutrons and of x-rays for medical therapies. Although the field of room-temperature fusion is littered with scandals and dubious discoveries, this device appears to be different: It has already won over some skeptics.

“My first reaction was, ‘Oh, God, not again,’” says Michael Saltmarsh, a physicist at Oak Ridge National Laboratory in Tennessee. “But upon reading the paper, I thought that it was really neat; it’s such a cute way of making an accelerator.”

In this week’s issue of *Nature*, Seth Putterman, a physicist at the University of California, Los Angeles, and colleagues describe the fusion device, which is about the size of a small bucket. At its heart is a little crystal of lithium tantalate—a material that has a peculiar property: It is pyroelectric.

Pyroelectricity is related to the better-known phenomenon

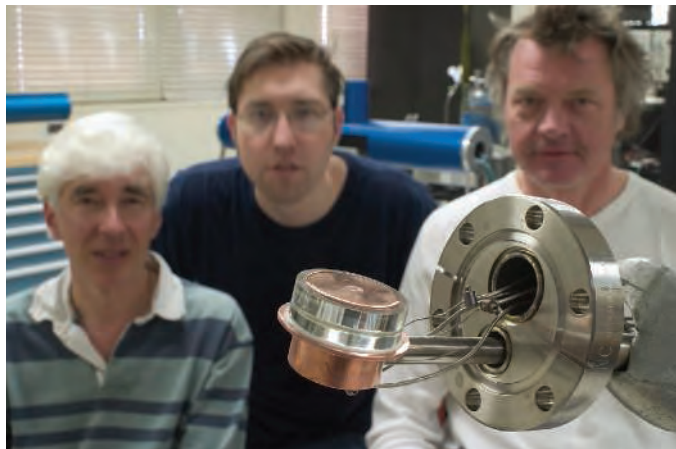
of piezoelectricity. If you squash a piezoelectric crystal, such as quartz, the electrons in the crystal rearrange themselves so that one side of the crystal becomes positively charged and the other negatively charged, creating a voltage difference between the two ends. A pyroelectric crystal does the same thing if you heat or cool it.

Putterman’s group cooled the pyroelectric lithium tantalate crystal and put it in a chamber full of deuterium gas. When they warmed the crystal with a heater, the pyroelectric effect created a huge electric field near a tungsten needle attached to the crystal. The crystal and needle essentially focused all the energy of the crystal’s heating to the very tip of the tungsten spike. When deuterium (atoms of heavy hydrogen, with a proton and a neutron in the nucleus) ventured near the tip, the field stripped off their electrons and shot the charged nuclei into a deuterium-loaded target. Some of those deuterium ions

struck deuterium in the target and fused, releasing protons, neutrons, and energy.

“Neutrons were everything to this experiment,” says Putterman, whose team spent 2 years developing a neutron detector for the experiment. “We can grab single neutrons—the actual trajectory of each neutron.” The data show about 900 neutrons per second flying away from the target with the energies one would expect from a fusion reaction. “If you look at the raw data, we maintain that it’s incontrovertible,” Putterman adds.

Saltmarsh, a neutron expert, says he is convinced but adds that the device is unlikely to be useful for generating energy. “Even if it had 100% efficiency, you can’t make net energy. The ion beam is slowing down in the



Small wonder. UCLA physicists Seth Putterman (left), Brian Naranjo, and Jim Gimzewski say their portable deuteron gun can fuse atoms.

target, and it loses energy,” more than counteracting the energy gained from fusion, he says. Saltmarsh adds that the device doesn’t produce enough neutrons yet to be commercially useful: “At this level [of intensity], it has curiosity value and lab value; it would make a good device for demonstrations. I wouldn’t mind having one in my lab.”

Putterman hopes a more refined device will produce a million or so neutrons a second. A hand-held neutron generator like that might have homeland-security applications, such as probing for fissile materials in sealed containers. Putterman says the device can also accelerate electrons into a target, producing x-rays. “A 1-millimeter crystal should be able to deliver therapeutic doses,” he says.

Whether or not the device proves useful, the idea of a simple fusion machine captivates physicists. “There [are] no moving parts,” marvels Saltmarsh. “Just heat it up.”

—CHARLES SEIFE

Congress Probes Charges of Harassment at NIH

Two congressional committees are looking into charges of sexual harassment at the National Institutes of Health (NIH). The complaints arose after National Institute of Allergy and Infectious Diseases (NIAID) staffer Jonathan Fishbein alleged that a landmark clinical trial, which found that the drug nevirapine can reduce mother-to-infant transmission of HIV, was seriously flawed.

An Institute of Medicine panel recently concluded that, although researchers failed to report some adverse events data, the NIAID-funded nevirapine trial was scientifically valid (*Science*, 15 April, p. 334). But the Senate finance committee is now following up on a complaint from Fishbein accusing a supervisor of sending profane e-mails, as well as recent depositions by two female NIAID staffers involved in monitoring the trial that allege inappropriate behavior by supervisors. The committee chair, Senator Charles Grassley (R-IA), has asked NIH for more information, citing Associated Press articles that first reported the depositions and evidence obtained by committee staff.

The matter is also under review by the House Energy and Commerce Committee, chaired by Joe Barton (R-TX). An NIH spokesperson says the agency is conducting its own investigation as well.

—JOCELYN KAISER

Two Israeli Universities Targeted for Boycott

CAMBRIDGE, UNITED KINGDOM—The U.K. Association of University Teachers (AUT) has called for a boycott of two Israeli universities said to be supporting Israel’s occupation of Palestinian territory.

After little debate, the group voted 22 April that its members—from professors to university support staff—should shun Bar Ilan and Haifa universities. The boycott’s proponents claim that Bar Ilan is affiliated with a West Bank school “in the illegal settlement of Ariel,” and that the University of Haifa has harassed a senior lecturer who guided a student’s investigation into the conduct of Israeli soldiers.

The universities deny the allegations. Moshe Kaveh, president of Bar Ilan University and a well-known physicist, called AUT’s decision “very unbalanced” and “shameful.” AUT, meanwhile, has asked members to delay implementing a boycott pending legal advice.

—ELIOT MARSHALL

Earth Observation Program 'At Risk,' Academy Warns

Senior U.S. scientists are urging NASA and the Bush Administration to reverse plans to postpone or cancel several satellites designed to gather data on the land, sea, and atmosphere. In an interim report* released this week, a National Research Council (NRC) panel warns that “the nation’s Earth observation program is at risk” from tight budgets at NASA and other federal agencies. Their advice would put the enterprise on a healthier track for the coming decade, they say.

The final report, due out in late 2006, will lay out a course for space-based Earth observation with clear priorities, similar to those in astronomy, planetary science, and solar and terrestrial physics. But NASA’s recent moves to scale back future programs and turn off currently operating satellites prompted committee members to push through a report that could influence congressional debate on the 2006 budget, which goes into effect on 1 October. Coincidentally, the interim report was released the same day a team of NASA and outside scientists met in Washington, D.C., to consider which of half a dozen currently operating Earth science satellites should be shut down.

The 18-member NRC panel was co-chaired by Richard Anthes, president of the University Corporation for Atmospheric Research in Boulder, Colorado, and Berrien Moore, a bio-

* *Earth Science and Applications from Space: Urgent Needs and Applications to Serve the Nation*, National Academy Press.

geochemist at the University of New Hampshire in Durham. Its report notes that several federal agencies supporting earth sciences research are under similar budget pressures. “Additional funds will be needed,” the panel concludes, although it gave no estimate.

The panel did not shy away from specific recommendations. NASA should proceed “immediately” with the oft-delayed Global Precipitation Measurement Mission, it concluded. The spacecraft, with contributions from Japan, would provide important data on Earth’s water cycle. In March, NASA’s chief of Earth observation, Mary Cleave, told a NASA panel that “we’re trying to hold on to a 2010 launch” using a Japanese rocket.

The NRC panel also wants NASA to resume work on the \$100 million Geostationary Imaging Fourier Transform Spectrometer that could improve detection of weather changes leading to tornadoes, floods, and

hurricanes. NASA, which is working with two universities and the National Oceanic and Atmospheric Administration (NOAA), canceled the mission in February. But the panel urges the agency to finish the instrument and seek international help in launching the satellite by 2008.

In addition, the interim report recommends “urgent reconsideration” of a planned cancellation of three other missions: a probe called Ocean Vector Winds to enhance the accuracy of severe storm forecasts, a spacecraft to continue Landsat observations, and the Glory satellite to measure atmospheric aerosols. In a proposed cost-saving move, the committee suggests that the instruments planned for the canceled missions could be flown instead on the National Polar-orbiting Operational Environmental Satellite System (NPOESS), which is being built for a 2010 launch.

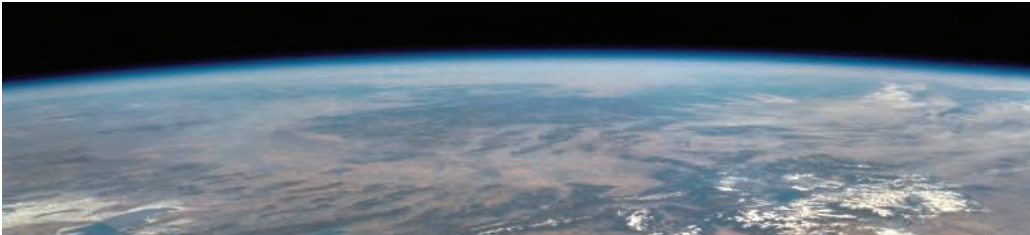
The panel wants NASA to resume Explorers, a program of small satellites now on hold, and launch one per year. Panel members also lament cuts to research and analysis funds used primarily by university researchers to analyze NASA satellite data. If NASA does not reverse the trend, the report states, “the long-term consequence will be a diminished ability to attract and retain students interested in using and developing Earth observations.” That drop-off, in turn, would “jeopardize U.S. leadership in both earth science and Earth observations.”

Shutting off existing NASA satellites, many earth scientists worry, could mark the start of a U.S. retreat on global data gathering. And White House science adviser John Marburger had few comforting words during an 18 April press conference touting a global system of Earth observation. “NASA just can’t keep putting money into continuing operations” of satellites beyond their expected lifetime, he said. Marburger blamed the confusion over how and when NOAA will inherit some responsibilities for gathering climate data on the recent change of leadership at NASA.

But money is also a key issue. NOAA chief Conrad Lautenbacher made it clear at the same ▶



Moore sees less. Berrien Moore hopes interim report can help reverse the decline of earth science.



Mission	Measurement	Societal Benefit	Status
Global Precipitation Measurement	Precipitation	Reduce vulnerability to floods and droughts; improve forecasts of hurricanes	<i>delayed</i>
Atmospheric Soundings From Geostationary Orbit	Temperature and water vapor	Improved weather forecasts and severe storm warnings	<i>canceled</i>
Ocean Vector Winds	Wind speed and direction near the ocean surface	Improved warnings to ships at sea; better predictions of El Niño	<i>canceled</i>
Landsat Data Continuity	Land cover	Monitor land-use changes; find mineral resources	<i>canceled</i>
Glory Instrument	Optical properties of aerosols; solar irradiance	Improved understanding of climate change	<i>canceled</i>
Wide Swath Ocean Altimeter (on the Ocean Surface Topography Mission)	Sea level in two dimensions	Monitor changes that affect fisheries, navigation, and ocean climate	<i>instrument canceled, descope of mission</i>

Missions impossible? NASA is delaying or canceling several long-planned earth science missions.

press conference that his cash-strapped agency will not accept a request from NASA to pay for operations of an existing spacecraft like the Tropical Rainfall Measuring Mission, which the space agency intends to shut down soon.

Earth scientists are looking for what Moore calls a “politically compelling agenda” to overcome such obstacles—and quickly. Congress is at work on NASA’s 2006 request,

and the agency already is preparing its 2007 wish list for the White House. “We’ve been running on the fumes of the past, and we need a vehicle to bring the community together,” says Moore. But he and his colleagues may have trouble finding the fuel they need—from Congress, the White House, and the agencies—to keep the United States at the forefront of earth science. —ANDREW LAWLER

DEPARTMENT OF ENERGY

Falling Budget Could Force Choice Between Nuclear Science Facilities

TAMPA, FLORIDA—A panel of experts weighing the future of nuclear physics in the United States may soon recommend shutting down a major Department of Energy (DOE) facility as a way to cope with a dismal budget.

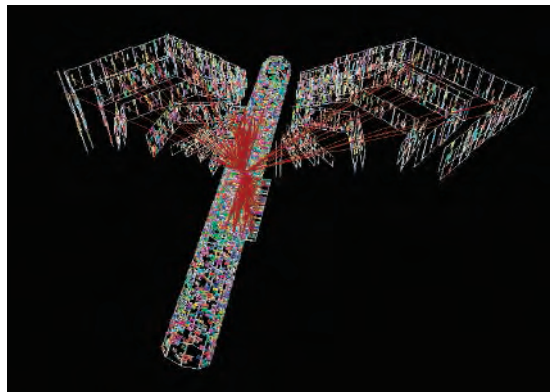
Last month, DOE and the National Science Foundation asked their Nuclear Sciences Advisory Committee (NSAC) to reevaluate the government’s long-term plans for nuclear physics. The trigger is the Bush Administration’s proposed 8.4% cut in DOE’s nuclear physics program for the 2006 budget year that begins on 1 October. Such a decrease, if adopted by Congress, would drastically reduce running times by as much as 60% at the two flagship nuclear physics experiments in the United States, CEBAF at the Thomas Jefferson National Laboratory (JLab) in Newport News, Virginia, and RHIC at Brookhaven National Laboratory in Upton, New York.

At a minimum, those cuts will mean layoffs and the shuttering of two of RHIC’s four experiments. But the big question for the NSAC panel is whether such tinkering will be enough. The language of the charge letter is quite ominous: “This funding level, projected into the outyears, is not sufficient . . . to continue operations of the program’s two major facilities, RHIC and CEBAF, as they are presently conducted.” And although DOE officials won’t prejudge the work of the panel, which was asked to make recommendations based on three budget scenarios, it’s clear that the stakes are high. “Looking at the magnitude of the problem, something is going to have to happen,” says Dennis Kovar, associate director for nuclear physics in DOE’s Office of Science. “To develop capabilities for the future, tough decisions have to be made.”

The panel, chaired by physicist Robert Tribble of Texas A&M University in College Station, must decide how to handle a shortfall that DOE officials estimate will grow to about \$130 million by fiscal year (FY) 2011. That amount is

roughly one-third the size of DOE’s current nuclear physics program. “What we’ve heard, consistently, is that if we let the program go on like [it is structured in] FY ’05, by FY 2011 it will be dead,” says Tribble. “I don’t think that’s an option.”

But physicists say that the idea of terminating either facility prematurely is also abhorrent. “Do you cut off the left hand, or do you cut off the right hand?” asks Gerald Miller, a nuclear



Big crunch. The Phobos experiment, which tracks colliding particles, will be shut down in response to a budget squeeze that could also claim the rest of RHIC.

physicist at the University of Washington, Seattle, whose theoretical work interprets data gathered at both JLab and RHIC. Another issue for the panel is that the nuclear physics programs at Brookhaven and JLab make up more than 50% and 96% of the labs’ income from DOE, respectively. So the death of an experiment could also determine the fate of the lab itself.

Whatever the subcommittee does, speed is essential. “It’s due at the end of June,” says Yale University’s Richard Casten, who chairs the parent NSAC. “[This report] will have a number of important implications, but there’s no time for a new long-range plan.”

Casten says it’s always possible that the budget situation might improve. But in the meantime, the nuclear physics community may soon learn which hand is on the chopping block. —CHARLES SEIFE

Griffin Moves Fast To Reshape NASA

NASA’s new chief Michael Griffin is making his mark after only 2 weeks on the job. Last week, Griffin reversed a decision by his predecessor Sean O’Keefe to split the NASA Advisory Council into two panels—a move that critics feared would weaken its ability to receive impartial external advice. He also told managers that congressional pork—or “earmarks”—would be funded expeditiously by NASA. O’Keefe had refused to dispense the money for those earmarks in the current budget.

Griffin decided to keep NASA interim chief Fred Gregory as his deputy but has created a position to handle the agency’s day-to-day operations. It will be filled temporarily by Courtney Stadd, an entrepreneur who has worked for several agency administrators. —ANDREW LAWLER

Netherlands Reports First vCJD Case

The Netherlands has become the fifth European country affected by variant Creutzfeldt-Jakob disease (vCJD), the human counterpart of bovine spongiform encephalopathy (BSE). On 21 April, health authorities reported that a 26-year-old woman had been diagnosed with the fatal brain affliction. The Netherlands has registered at least 77 cases of BSE.

Of 171 cases of vCJD so far, 155 have occurred in the United Kingdom, nine in France, two in Italy, and one in Ireland. Four additional patients from outside Europe had all lived in the U.K. for varying periods. —MARTIN ENSERINK

UC Retains Oversight of Lawrence Berkeley

It came as little surprise, but last week the U.S. Department of Energy (DOE) awarded the University of California (UC) a 5-year contract to manage Lawrence Berkeley National Laboratory (LBNL). UC has run LBNL since its inception more than 60 years ago, but this was the first time the university had been asked to submit a competitive bid. UC was reportedly the only bidder for the contract, valued at \$2.3 billion.

UC President Robert Dynes says the university is still considering whether to submit a bid to continue managing Los Alamos National Lab (LANL) in New Mexico under a new contract that begins 1 October. It will do so, Dynes says, if DOE puts the emphasis in LANL’s mission on science and technology instead of weapons development. Defense giants Lockheed Martin and Northrop Grumman also plan to bid for the LANL contract. —ROBERT F. SERVICE

CREDIT: BROOKHAVEN NATIONAL LABORATORY

SCIENCE POLICY

Marburger Asks Social Scientists for A Helping Hand in Interpreting Data

Will the growing number of engineers graduating from Chinese universities be a boon or bane to the United States and the rest of the world?

John Marburger would like to tell his boss, President George W. Bush, how that trend might affect the U.S. technical workforce and the country's economy—or even how long it's likely to persist. But the president's science adviser says he'd be flying by the seat of his pants. "I won't take a position on whether it's good or bad based on the data," says Marburger, "because we don't have adequate models."

Last week Marburger challenged the scientific community to help him find answers to a host of questions like these that puzzle science policymakers. "I am suggesting that the nascent field of the social science of science policy needs to grow up, and quickly," Marburger told a Washington, D.C., gathering sponsored by AAAS (which publishes *Science*). Economists have applied "behavioristic" tools successfully in other fields, says Marburger, pointing to analyses of how changes in retirement patterns might affect Social Security. He urged scientists to incorporate "the methods and literature of the relevant social science disciplines" to explore trends such as the community's "voracious appetite" for federal research funding, the "huge fluctuations" in state support for public universities, and the continuing advances in information technology.

Marburger's call to statistical arms was generally welcomed by policy analysts, who agreed that their field hadn't made much progress on the big questions confronting decision makers. "We operate with blinders on," says Daniel Sarewitz of Arizona State University in Tempe, a former congressional staffer who studies the interplay of science and society. "Rather than simply tracking the growth in industrial R&D, for example, we also need to look at how that affects public sector investment. The set of assumptions that goes into S&T policy is unbelievably oversimplified."

That lack of rigor, speculates Harvard economist Joshua Lerner, part of a group studying U.S. innovation policy, could be a result of the limited interaction between the disciplines. "A lot of science policy has an amateur-hour flavor to it because it's done by scientists who aren't familiar with the principles of the social sciences," he says. "But it's also our fault. We economists haven't communicated as well with other disciplines as we should."

Another factor is the sheer difficulty of



Supermodel. U.S. science adviser John Marburger wants better econometric models of research trends.

coming up with a theoretical framework that takes into account enough of the important variables to generate useful results. "Such a model has proved to be elusive," says Rolf Lehming, who oversees the National Science Foundation's biennial volume: *Science and Engineering Indicators*. Previous efforts to nurture such a community of scholars were abandoned, notes Mary Ellen Mogege, a science policy analyst at SRI International in Arlington, Virginia, including the 1995 elimination of the congressional Office of Technology Assessment.

Marburger says that he believes a new effort can be mounted at minimal cost. "We're not talking about a lot of money; ... funding is not a rate-limiting factor in this equation." But others see a federal role as crucial. Connie Citro, who directs the National Academies' Committee on National Statistics, says that "there needs to be at least a signal [from the federal government] that proposals would be welcome." Sarewitz admits that a plea for federal support is self-serving, but he adds, "that's what drives academics in any field."

—JEFFREY MERVIS

U.S. PUBLIC SECTOR

Agency Kills New Performance Rules

A plan by a U.S. government agency to reward or punish its scientists based on their ability to drum up paying customers has been withdrawn after a watchdog group complained that it would make the researchers "sing for their supper."

The plan would have affected some 30 scientists at two Denver, Colorado-based divisions of the U.S. Bureau of Reclamation working on a broad range of environmental assessments required under federal laws to safeguard ecosystems and their inhabitants. The idea was to link scientists' annual performance evaluations to the amount of business they generated, akin to rating a lawyer's prowess at racking up billable hours. Based on a five-point scale, "exceptional" employees would haul in over \$529,000—more than three times what they cost the govern-

ment in annual salary and benefits. A mere \$150,000 or so would be deemed "minimally successful," which in federalese is tantamount to loafing on the job.

That metric, put in place earlier this year by two managers within the bureau's Technical Services Center, triggered squawks from employees who thought public servants should not be judged on how well they peddle their expertise. On 20 April the Washington, D.C.-based Public Employees for Environmental Responsibility (PEER) issued a press release decrying the idea of monetary quotas and warning that scientists might feel pressured to tweak a report to keep the customer happy. "They're worried about these new rules," explained PEER program director Rebecca Roose. "But they didn't know how to fight them."

The answer, apparently, was to go public. Two days later, the bureau withdrew the new evaluation system, which replaced what bureau spokesperson Trudy Harlow called a simple "pass/fail system" for judging an employee's performance. "We became aware that some scientists were unhappy with it and that there was a perception it could taint the quality of our service," says Harlow. "We would never want that to happen." She said that although the center is a fee-for-service operation within the Department of the Interior, all its customers are public agencies and "we don't compete with the private sector."

PEER is pleased with the bureau's decision, says Roose, but it plans to monitor the situation in case such a quota system reappears in another guise.

—JEFFREY MERVIS

As two vaccines against a sexually transmitted virus approach the market, public health experts are debating who should receive them—women, boys, or girls—and how to make them affordable in developing countries where the need is highest

High Hopes and Dilemmas for a Cervical Cancer Vaccine

Investigators who stage large, placebo-controlled studies go into them with a great deal of trepidation. It is make-or-break time for vaccines or drugs that have consumed years of their labor—not to mention many millions of dollars. All too often, exciting results hinted at in animal and limited human tests don't pan out. Sometimes, devastating side effects surface. Even when the trial is a success, the naked data that emerge frequently contain unsightly blemishes. But for researchers who developed two different vaccines against human papillomavirus (HPV), the results from clinical trials so far have generated little angst. Tested in more than 3000 participants, the vaccines have shown stunning, and nearly identical, curves: Both prevented persistent infection with this widespread, cancer-causing virus in a whopping 100% of the vaccinated women and reduced cervical abnormalities by more than 90%. "We're pinching ourselves," says John Schiller, a papillomavirus researcher at the U.S. National Cancer Institute (NCI) in Bethesda, Maryland, whose lab helped developed a key technology used to make both vaccines. "It's better than we could have imag-

ined." Yet these attractive, early results have also pushed to the fore vexing questions that, ultimately, will affect how much disease and death the vaccines prevent.

The two vaccines—made by Merck & Co. of Rahway, New Jersey, and GlaxoSmithKline (GSK) Biologicals of Rixensart, Belgium—must still prove safe and effective in phase III efficacy trials now under way in more than 50,000 people in several countries (see table, p. 621). But Merck has announced that it plans to file for approval with the U.S. Food and Drug Administration (FDA) before the end of the year, and GSK says it will seek approval in Europe and other unspecified countries in 2006. "The fact that we've done this as fast as we have is remarkable," says Diane Harper, a clinician at Dartmouth Medical School in Lebanon, New Hampshire, who has worked on trials of both vaccines. Harper, Schiller, and many other researchers expect that, barring any big surprises, both vaccines will make it to market with relative ease.

In anticipation, the companies, public health officials, clinicians, researchers, and even the public itself have already started to ask who, exactly, should get the vaccines

first: Adolescent girls? Older women? Boys and men? How long will vaccine protection last? Will developing countries, which account for 80% of the deaths from cervical cancer, have to wait years before they get the products? How will the vaccines affect the tests that developed countries routinely use—with great success—to screen for cervical cancer? For that matter, how much will the vaccines actually alter cancer rates in the wealthy world? And how will these issues affect vaccine sales?

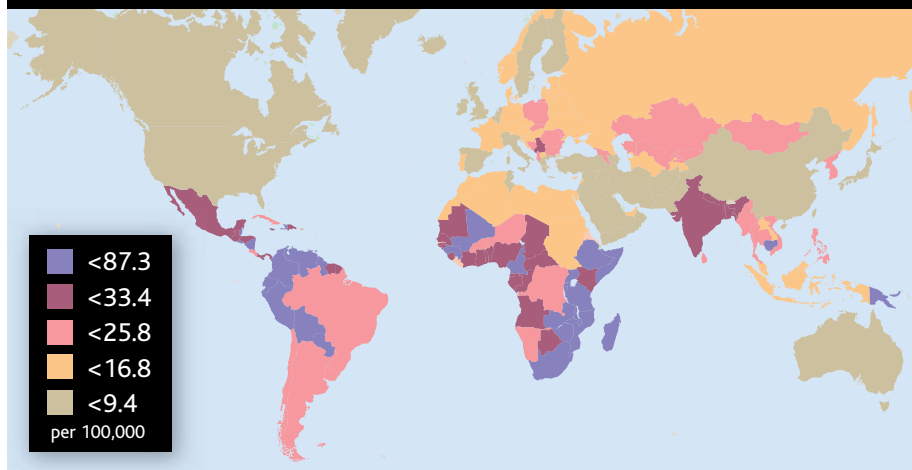
Many of these critical questions will be front and center this week at the 22nd Annual International Papillomavirus Conference and Clinical Workshop in Vancouver, Canada. "The issue is now very hot," says F. Xavier Bosch, an epidemiologist who has contributed to studies of both vaccines and works at the University of Barcelona's Catalan Institute of Oncology. Firm answers, however, will likely remain few and far between for some time to come.

Rapid evolution

In 1975, virologist Harald zur Hausen presented provocative evidence that HPV, a common infection spread through skin-to-skin contact and sex that was believed to lead to serious disease only rarely, could cause cervical cancer. Zur Hausen, who for 20 years headed the German Cancer Research Center in Heidelberg, led a team that by the early 1980s had isolated several genotypes of the virus, some of which they linked to genital warts and others to cervical cancer. "For quite a while, we faced a lot of resistance," says zur Hausen, now a professor emeritus. But as the polymerase chain reaction assay improved the ability to detect viral DNA, epidemiological data accumulated that backed zur Hausen's theories. Indeed, one 1999 report found HPV DNA in 99.7% of cervical cancers studied, conclusive evidence that persistent infection with the virus causes the disease.

Nearly half a million women worldwide developed cervical cancer in 2002 (see map, p. 618), and it killed 270,000, accord-

Cervical Cancer Rates Worldwide



Disproportionate impact. As the Pap smear has become common in wealthy countries, cervical cancer cases and deaths have become increasingly concentrated in the poorer areas of the world.

ing to the latest data from the International Agency for Research on Cancer (IARC). In developed countries, use of the Papanicolaou test, or Pap smear—which swabs the cervix and looks for abnormal cells—has dramatically cut cervical cancer rates over the past 50 years: Only 5000 American women died from the disease in 2002, a 75% drop in mortality since 1950. But much of the world still does not routinely use the Pap smear, making the need for a vaccine that much more pressing.

Scientists have identified more than 100 genotypes of HPV, only 40 of which infect the genital tract; of these, about 15 put women at “high risk” for cervical cancer. In the vast majority of cases, the immune system clears HPV infections before they can cause harm.

Bosch helped conduct an IARC-coordinated study published last year in the *International Journal of Cancer* that examined the HPV types detected in more than 3000 women from 25 countries who had cervical cancer. The researchers found relatively modest geographical differences, with two types, HPV 16 and 18, occurring in more than 70% of the cases. The next five most prevalent types together accounted for 20% of the cases (see figure, p. 621).

Both Merck and GSK used HPV 16 and 18 as the backbones of their vaccines and also relied on the same basic technology. In the early 1990s, studies done by NCI’s Schiller and Douglas Lowy and a handful of other groups (who remain mired in patent disputes that GSK and Merck have settled through a cross-licensing agreement) showed that stitching the gene for HPV’s L1 protein into a different virus or yeast led to the self-assembly of viruslike particles. “That was the major breakthrough,” says virologist Gary Dubin, a vice president for clinical development at GSK. These empty shells of L1 contain none of HPV’s cancer-causing DNA (see sidebar) and mimic HPV’s shape; this suggested that they would safely trigger effective immune responses if injected into people. The viruslike particles could also be produced in high quantities, circumventing a formidable roadblock to vaccine manufacturing: HPV grows poorly in lab cultures.

The two vaccines do have marked differences. Merck has included two additional genotypes, HPV 6 and 11, which cause genital warts in both sexes. Merck added these two types in part to create an incentive for males to receive the vaccine; vaccinated males, in turn, might reduce viral spread to women. “Men are very worried about genital warts because they’re highly visible,” explains Eliav Barr, head of Merck’s HPV vaccine clinical trials program. “Why in the world would a young adult male or an adolescent male want to get

HPV’s Peculiarities, From Infection to Disease

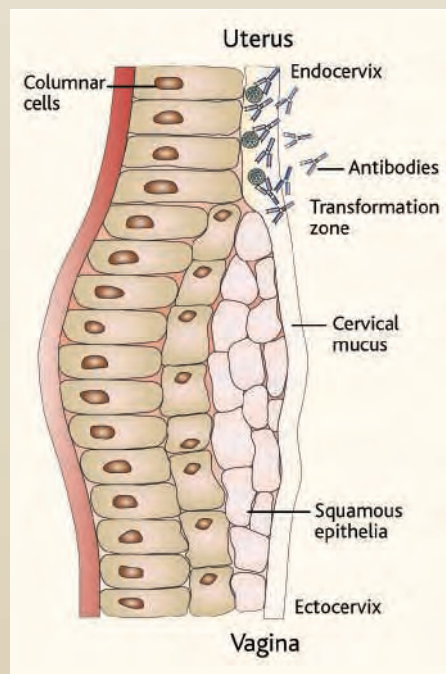
Human papillomavirus (HPV) is an odd bug. Not only does it come in more than 100 different varieties that infect different cells, but fewer than half of those infect the genitalia; some cause cancer, some cause warts, and some don’t seem to do much of anything. Whereas most sexually transmitted infections thrive within the body’s blood or nerve cells or internal organs, HPV resides in the skin, notes Diane Harper, a clinician at Dartmouth Medical School in Lebanon, New Hampshire, who tests HPV vaccines. “It does not invade the body,” says Harper. “It lives in the wrapping paper that surrounds the present. It doesn’t infect the present.”

In the cervix, HPV infects the epithelial cells that lie just under the mucosal surface. The viral types most responsible for causing cervical cancer, such as HPV 16 and 18, make proteins that powerfully bind two tumor suppressors, known as p53 and retinoblastoma protein. Blocking these tumor suppressors allows the squamous epithelial cells to divide abnormally, and cancer occurs for unknown reasons when they meet with columnlike columnar cells in what’s known as the transformation zone (see illustration).

The vaccines that have moved furthest in clinical trials contain a viral protein called L1, which forms the bulk of HPV’s outer shell. Injecting L1 into muscles triggers production of antibodies in the bloodstream, which then “transudate,” or pass into, the basement membrane of the cervix and up to its mucosal surface. If HPV shows up, the L1 antibodies presumably bind the protein and block HPV from establishing an infection.

In both vaccinated and unvaccinated women, if the virus dodges the initial immune response and wangles its way into the epithelium, immune cells that specifically eliminate infected cells, combined with a continued antibody assault, typically clear the infection. But when the attack on HPV fails, the virus can live in the body for many years, impervious to these types of preventive vaccines. And the longer HPV sticks around, the more chances it has to cause a life-threatening cervical cancer.

—J.C.



Blocking entry. Antibodies triggered by the vaccine presumably bind the L1 protein and prevent HPV infection.

vaccinated with a vaccine that would not in general help him out?” The vaccines also have different immune-boosting agents called adjuvants. Merck formulates its HPV with aluminum, the only adjuvant used in FDA-approved vaccines. GSK uses AS04, a proprietary adjuvant that contains aluminum and a bacterial lipid. Europe already has approved vaccines containing AS04.

When it comes to efficacy, the phase II studies published to date have remarkably similar results. Because it can take a decade or more for HPV to cause cervical cancer, the vaccine trials rely on easier-to-measure endpoints that are linked to the disease, including a cellular abnormality called cervical intraepithelial neoplasia (CIN) and infection with the virus itself. Data came

first from a multicenter study of Merck’s original formulation, which contained only HPV 16. Published in the 21 November 2002 *New England Journal of Medicine*, the study in 1500 women between 16 and 23 years of age found that all of the 41 participants who had “persistent” HPV 16 infections—two detections within 4 months—had received a placebo shot, meaning the vaccine offered 100% protection. The nine cases of HPV 16-related CIN all occurred in placebo recipients, too. “It doesn’t take much of an immune response to clear HPV infections,” concludes Laura Koutsky, an epidemiologist at the University of Washington (UW), Seattle, who was the first author of the study.

Next, researchers reported in the 13

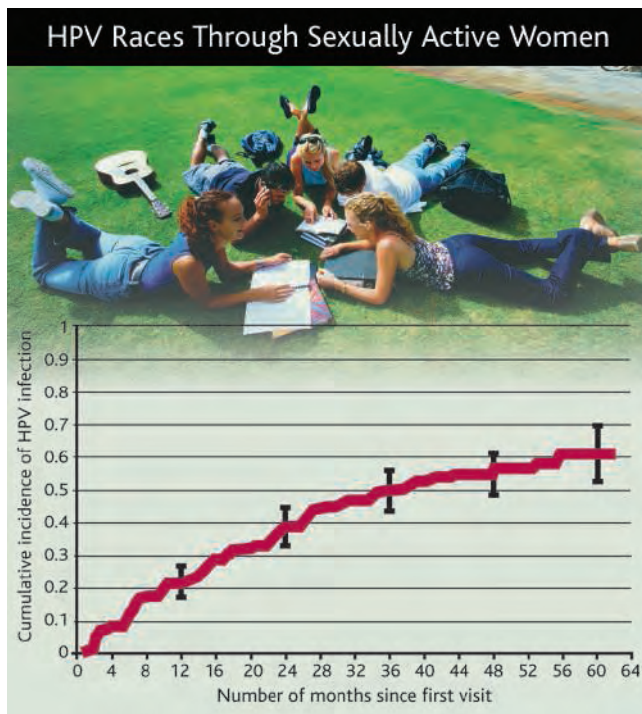
November 2004 issue of *The Lancet* that GSK's HPV 16/18 vaccine conferred 100% protection against persistent infection with those types in a placebo-controlled study that involved 700 women aged 15 to 25. CIN occurred in six placebo recipients and one vaccinated woman who had evidence of a persistent infection with a high-risk HPV type not in the vaccine. Then on 7 April 2005, *Lancet Oncology* published results online from a study of Merck's quadravalent vaccine in 500 women. Although the numbers were smaller, the vaccine achieved 89% protection against persistent infection and completely prevented CIN and genital warts. "It's very interesting that two vaccine candidates that have been produced independently and run through clinical trials in very independent ways show the same results," says Sonia Pagliusi, who heads the HPV vaccine project for the World Health Organization (WHO) in Geneva, Switzerland.

"I am rather surprised and enthusiastic about the similarities, and I hope the dissimilarities are details."

Who goes first?

Merck launched phase III efficacy trials in December 2001; GSK started its pivotal licensure studies in mid-2004. Both companies will need stricter evidence of efficacy before winning regulatory approval; specifically, they must show protection from advanced stages of CIN, known as 2 and 3, which have more definitive ties to cervical cancer and on average develop within about 3 years of infection. But given the phase II data and the possibility that an HPV vaccine could come to market next year, WHO just 2 weeks ago held a meeting with leading vaccine experts to discuss steps for introducing the vaccines to developing countries. Similarly, the Advisory Committee on Immunization Practices (ACIP), which helps steer U.S. vaccine policy, held its first powwow on the potential use of the vaccine in February. "We anticipate being on the ACIP agenda every meeting until the vaccine is licensed," says Lauri Markowitz, an epidemiologist with the U.S. Centers for Disease Control and Prevention in Atlanta, Georgia, who coordinates an ACIP working group on HPV vaccines.

One of the trickiest questions ACIP will have to address is the age group that should receive the vaccine. As a provocative study by UW's Koutsky and her colleagues showed, HPV—which can spread even when condoms



Big virus on campus. A University of Washington study found that more than 60% of college women became infected over 5 years.

are used—races through a population of young women soon after they become sexually active. Every 4 months, Koutsky's group tested for HPV in 18- to 20-year-old college students who initially were negative for the virus. Five years into the study, more than 60% of the nearly 300 women at some point had become infected with HPV. This leads Koutsky and many others to conclude that the vaccine ideally should be given to girls who are between 9 and 12 to protect them before they become sexually active. Already, some religious groups in the United States have voiced strong reservations, as they worry that vaccinating young girls will give them a green light to have sex. Koutsky balks at this. "Why don't you think of this as a red light for cancer?" she asks.

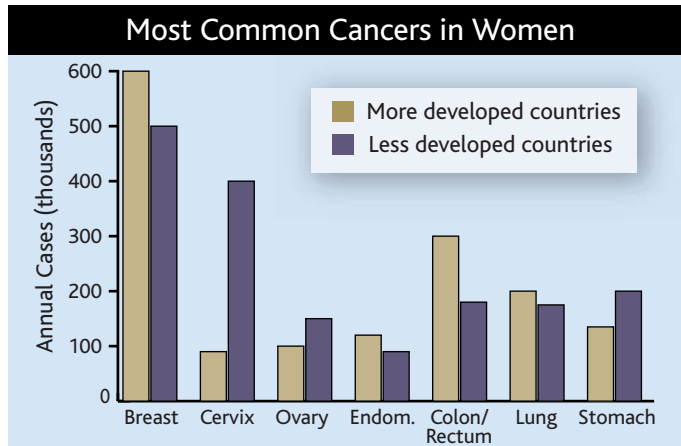
King Holmes, a sexually transmitted infection (STI) expert at UW, says HPV vaccine proponents must strive to reach a consensus with the concerned parents. "You can protect a woman against HPV in more than one way: One is to avoid risky sex and the other is a vaccine," says Holmes. And he thinks it helps to emphasize that HPV is the most ubiquitous STI. "HPV is really unlike any of the other

sexually transmitted pathogens," says Holmes. "You don't have to have a lot of partners." That makes a vaccine doubly important.

Both Koutsky and Harper say it may work better to target late teens and young women first. "That would make perfect sense for the introduction, to make people feel better about it," says Koutsky. Harper notes that this would also cater to the group most interested in the vaccine. "I have 50 women over the age of 25 who will be outside my door waiting to get the vaccine," says Harper. "I don't see mothers lining up with their daughters and sons the day the vaccine is available." Public health campaigns face a new challenge, too: They typically have focused on vaccinating young children and the elderly, rarely targeting adolescents and young adults.

Scientific issues will also drive decisions about who should get the vaccine. Both Merck and GSK have small "bridging" studies ongoing in younger girls that will evaluate safety and immune responses. And data will have to address how long vaccine-induced immunity lasts: It of course doesn't make sense to vaccinate 9-year-olds if protection disappears after 3 years.

As for men, Merck 6 months ago launched an efficacy study that will assess the vaccine's ability to prevent penile infection, warts, and anal intraepithelial neoplasia. Margaret Stanley, an HPV vaccine researcher at the University of Cambridge, U.K., warns that the same product could work differently in men and women. She points to a recent trial of a preventive herpes vaccine made by GSK that failed in men but, in one subgroup of women, worked more than 70% of the time. "We're all very cautious, especially after the herpes vaccine result, about differences in protection in the genital tracts of men and women," Stanley says.



Underdeveloped?

Like many of her colleagues, Stanley has deep concerns that even if an HPV vaccine proves safe and effective, several years might pass before people in poor countries have access to it. "It's completely unacceptable if the vaccine works and the people who need it most don't get it," says NCI's Schiller, adding that India alone has 30% of the world's deaths from cervical cancer.


Both Merck and GSK say they will offer the vaccine at a discount to poor countries. Schiller worries that this trickle-down scheme will take too long. "We have to do this sooner rather than later," says Schiller. "We can't just wait to see what the big pharmas are going to do." And Stanley says she's concerned that neither company has aggressively moved to stage studies in developing countries to make sure that other infections common in those locales don't interfere with vaccine efficacy.

Schiller, who recently met with scientists in India to discuss HPV vaccine particulars, says scientists there and in China may well make versions of the vaccine themselves. "It's naïve to think that those people in those countries can't do everything we can," Schiller says. "And it's more likely to get to women faster if they make it in their own country." As for patents, both countries could potentially sidestep them, as they have done with some anti-HIV drugs. Zur Hausen also suggests that traditional recombinant proteins might prove as effective as the more-difficult-to-manufacture viruslike particles. (The GSK and Merck efficacy trials may well reveal a correlate of protection, such as antibody levels, that makes it vastly simpler to evaluate the efficacy of future generation vaccines.)

WHO, the Bill and Melinda Gates Foundation, IARC, and the Seattle-based Program for Appropriate Technology in Health (PATH) all say they want to grease the wheels that move vaccines from rich to poor. But so far, no vehicle exists. "The vaccine itself has moved a lot more quickly than many of us expected a few years ago," says Jacqueline Sherris of PATH, which has a program to increase cervical cancer screening in resource-limited settings. "That said, there's a flurry of activity now."

Limits?

For poorer countries, the notion that a safe and effective HPV vaccine has a downside is irrelevant. But for the developed world, researchers already have begun thinking about the limitations of the current Merck and GSK vaccines. Foremost among them: the number of HPV types they include that target cervical cancer. Vaccines that contain HPV 16 and 18 combined, after all, don't protect against roughly 30% of cervical cancer cases. As the massive international study done by Bosch and colleagues found, adding HPV 45 and 31 captures another 10% of cases, and the next

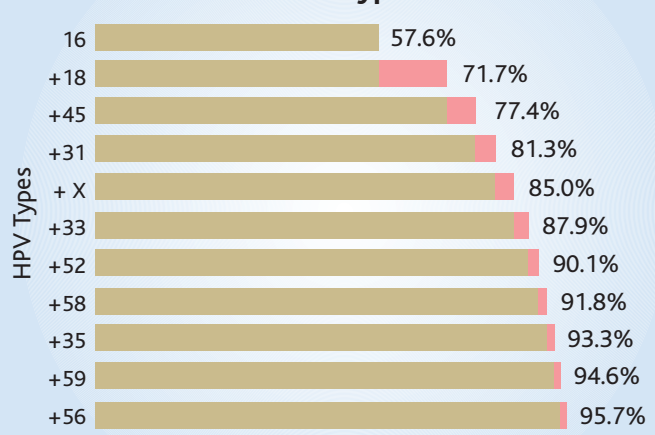
HPV Vaccine Efficacy Trials				
Manufacturer	Vaccine	Location	Participants	Projected End
 Merck	VLPs of L1 protein from HPV 6/11/16/18, made in yeast, aluminum adjuvant	U.S., S. America, Europe	17,800 women, 16 to 26 years old	2007
		U.S., S. America, Europe, Asia	3800 women, 24 to 45 years old	2008
		U.S., S. America, Europe, Asia, Africa	3700 men, 16 to 24 years old	2008
GSK	VLPs of L1 protein from HPV 16/18, made in baculovirus, AS04 adjuvant	U.S., S. America, Europe, Asia Pacific	18,000 women, 15 to 25 years old	2010
		Costa Rica (run by NCI)	12,000 women, 18 to 25 years old	2010

two most common types add another 5% (see table, above). But from there, individual types only add about 1% each. In the future, Bosch says he'd like to see a vaccine with four to six of the most common cancer-causing types of HPV. "Adding more, the benefit would be tiny," he says. Although no evidence exists that different genotypes can interfere with each other, both Merck and GSK note that adding types obviously creates more manufacturing difficulties and costs.

In countries that widely use Pap smears and other screens, Bosch and others say the current vaccines may have little impact on cervical cancer rates. "We might never see any effect on cervical cancer," says Bosch. Typically, it's women in lower socioeconomic classes who have the most cases of invasive disease in these countries, he explains, because they are the least likely to receive screens. "Chances are those same women will also escape vaccination," he says. And the analysts who weigh costs and benefits will surely assess how much bang the vaccines give for the buck.

The introduction of the vaccines could also have a negative impact on screening. Vaccinated women may wrongly think they no longer need regular Pap smears.

Global Prevalence of HPV Types in Cervical Cancer



Typical types. An international ranking of HPV types that put women at high risk of cervical cancer shows that the six most common ones account for nearly 90% of the cases. The Merck and GSK vaccines, now in efficacy trials, both contain HPV 16 and 18, the two most responsible for causing cervical cancer.

But the benefits of an effective vaccine clearly outweigh these concerns. In wealthy countries, fewer women will have abnormal screens in the first place, which means less anxiety, fewer cervical biopsies, and a reduction in the overtreatment that Bosch says now occurs. And if future generations of vaccines contain more HPV types, they will promise to cut cervical cancer rates more effectively than the best screens now available.

So although hopes are running high that the phase III trials will mirror the extraordinary data from the earlier studies, the complexity of further thwarting HPV in rich and poor countries alike has forced researchers to confront the naked truth: Having a safe and effective HPV vaccine is just a start.

—JON COHEN

Europe Steps Into the Open With Plans for Electronic Archives

In a flurry of new proposals, institutes and funding agencies are laying the groundwork for the free release of peer-reviewed papers

BERLIN AND PARIS—While moves in the United States to make scientific research results available—for free—at the click of a mouse have generated intense debate, European research organizations have quietly been forging ahead. Slowly but surely, they are starting to build and connect institutional and even nationwide public archives that will, according to proponents, be the megalibraries of the future, allowing anyone with an Internet connection to access papers that result from publicly funded research. “The cutting edge of the Open Access movement is now in Europe,” says Peter Suber of Public Knowledge, an advocacy group in Washington, D.C.

Institutes in Europe don't feel the intense heat from patient organizations, which helped drive the free-access movement in the United States. But many agree with its philosophy. Some say open archives offer research managers and funders a way to monitor scientific output; they can also increase access to dissertations, reports, and other “gray literature” that doesn't make it into journals. In many cases, they are out ahead of their own researchers, who, far from clamoring for open access, tend to ignore such archives unless they are required to deposit their own papers.

London's Wellcome Trust, for example, has taken one of the strongest public-access positions worldwide. The U.K.'s largest funder of biomedical research is planning to launch a system that will archive all papers produced by its grantees. Wellcome will require researchers to deposit a copy of the accepted manuscript within 6 months of publication. That goes much further than the U.S. National Institutes of Health (NIH) in Bethesda, Maryland, which decided to “strongly encourage,” but not require, grant recipients to post their papers in the U.S. National Library of Medicine's PubMed Central within 12 months of publication—a policy that has drawn heated opposition from some scientific societies and publishers who fear it will put some journals out of business (*Science*, 11 February, p. 825).

In the coming weeks, Wellcome plans to issue a call for applications to host the archive, which will be connected to PubMed Central. “We will be providing a door in England to the worldwide library,” says Robert

Terry, a senior policy adviser at the trust. Although the data behind the screen will be the same, the U.K. site will be tailored to U.K. users, he says, providing links to grant numbers so that users—especially funders—can track specific projects. To nudge researchers along, Terry says, the trust may consider an applicant's depositing record in decisions on future grants. Wellcome hopes to identify a host by early fall and have the database up and running early next year.

The U.K. Medical Research Council (MRC), the Biotechnology and Biological Sciences Research Council, the Department of Health, Cancer Research UK, and the British Heart Foundation are considering joining the project, which based on NIH's figures will likely cost at least \$1.5 million. “We are certainly very interested in what Wellcome is doing,” says Anthony Peatfield of the MRC. The seven U.K. Research Councils plan to announce their own public-access policy next month, which is expected to ask grant recipients to deposit their papers in an archive maintained either by their own institution or, if available, a centralized one like U.K. PubMed Central.

Similar projects are under way in France, Germany, and the Netherlands. The continent's open-access advocates got a boost in October 2003, when members of several of Europe's leading scientific organizations signed the so-called Berlin Declaration. It says that authors should retain rights to their papers—including the right to distribute electronic copies freely—and that all papers should be deposited in a public archive “maintained by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, interoperability, and long-term archiving.” So far, 56 organizations from 17 countries have signed the declaration, and many are starting to put it into practice. Publishers are concerned, says Sally Morris, executive director of the Association of Learned and

Professional Society Publishers, based in Clapham, U.K. For smaller journals in slower moving fields, free access, even with a 12-month delay, “could mean serious loss of subscriptions and journals collapsing,” she says. “The potential to destroy the journals that the open-access movement is parasitizing is very real indeed.”

In France, the government's four major research institutes—which together spend some €3.5 billion on research annually—6 weeks ago jointly declared their intention to move toward open archives. Furthest along is the National Center for Scientific Research (CNRS), which plans to expand an archive for physics and math papers that it has operated for 4 years. Eventually, the quartet may create a common database and a Web portal that archives as much of French research as possible, says Odile Hologne of the National Institute of Agricultural Research.



Old model. Proponents of open electronic archives say they are working to create the megalibraries of the future.

Ideally, the full text of all published papers would be archived, says Christian Bréchet, director-general of the Institute for Health and Medical Research (INSERM). But INSERM doesn't plan to force researchers to publish only in journals that accept this, Bréchet says, so for the time being, there will be gaps. “We have to be realistic,” he says.

Meanwhile, all 13 universities in the Netherlands have joined with the Netherlands Organization for Scientific Research (NWO), a major science funder, and the Royal Netherlands Academy of Arts and Sciences and the Royal Library to develop a network of databases called Digital Academic Repositories (DARE). Whether or not researchers will be obliged to participate is for each institute to decide, says program manager Leo Waaijers. But to pique interest and get the ball rolling, DARE will showcase the works of some 200 of the country's top scientists next month in a project dubbed

“Cream of Science.” Unlike the British agencies, however, NWO has no plans to use its muscle to enforce participation.

The German national science funding organization, the DFG, is also a signatory to the Berlin Declaration. It covers researchers’ expenses if they want to submit to open-access journals that require a publication fee. Spokesperson Eva-Maria Streier says the organization is considering strengthening its position by adding a clause to its grants that would require researchers to deposit papers in an institutional archive within a year of publication.

The experience of Germany’s Max Planck Society, which took a lead role in drafting the Berlin Declaration and hosted

the meeting where it was launched, highlights a few potential pitfalls. The organization has built a pilot archive, called eDoc, available to all Max Planck researchers. But participation is voluntary—and far from complete. Indeed, the Max Planck’s independent structure prohibits the society from requiring its researchers to archive their work. In addition, Max Planck officials have found that their historians, lawyers, biologists, and physicists have very different ideas about open access.

Indeed, leaders of several open-access initiatives note that their biggest challenge is not publishers’ restrictions on copyright but researchers’ inertia. Different tactics are being considered to overcome it. Terry says

he hopes the Wellcome Trust’s moves will help change that. “I describe it as passive resistance,” he says. He points to a study by the U.K.’s Joint Information Systems Committee that showed nearly 80% of scientists said they would deposit their papers in an archive if their funder required it. Only 5% said they would refuse. In France, researchers may be compelled to join by making only papers deposited in open archives count during their periodic evaluations, says CNRS’s Laurent Romary. Kurt Melhorn of the Max Planck Institute for Informatics in Saarbrücken and a leader of the eDoc project, says he hopes peer pressure will eventually do the trick: “It’s a question of critical mass.”

—GRETCHEN VOGEL AND MARTIN ENSERINK

Developmental Biology

Combing Over the Polycomb Group Proteins

From flies to people, the protein called Polycomb and its partners turn off genes and even an entire chromosome during development. They may also play a role in cancer

Scientists have known for decades that a gene called *Polycomb* plays a key role in establishing the body plans of organisms from fruit flies to humans. Exactly how it does this has been a big mystery, but recently that mystery has begun to yield.

The proteins produced by *Polycomb* and other genes with similar developmental effects—they’re called the Polycomb group proteins—for the most part turn off other

Kingston of Harvard’s Massachusetts General Hospital in Boston. “That’s been fascinating to a lot of us for years.”

The new work shows that the Polycomb group proteins, working in various combinations with one another, accomplish this feat by altering chromatin, the complex of DNA and associated histone proteins that together comprise a cell’s chromosomes. One set of the proteins first marks the genes to be silenced by attaching methyl groups to a specific histone called H3. A second set then

of the two X chromosomes carried by female cells, which is needed to prevent an overdose of X gene expression. In addition, mirroring findings on other key developmental control genes, researchers have recently linked abnormal expression of one of the *Polycomb* group genes to the development of prostate, breast, and other cancers.

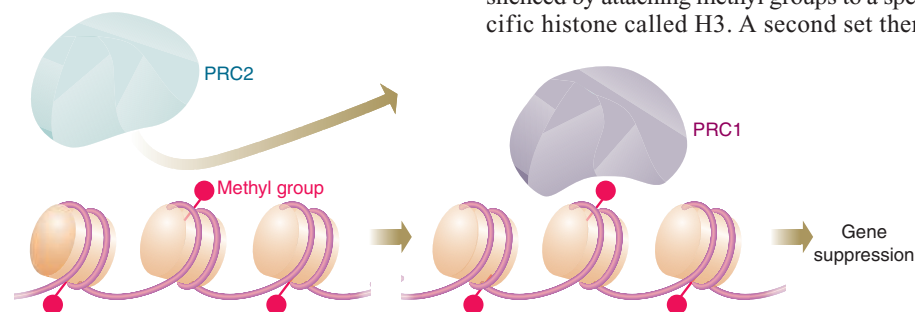
A venerable history

The *Polycomb* gene turned up nearly 60 years ago, discovered in experiments performed on fruit flies by Pamela Lewis, wife of the late Ed Lewis, a Nobel Prize-winning geneticist at the California Institute of Technology in Pasadena. Normal male fruit flies have bristly structures called sex combs on their front legs that they use for grasping females. Pamela Lewis identified mutant flies that also had sex combs on the second and third pairs of legs, hence the name Polycomb.

The development of the flies had apparently been altered so that their more posterior segments were producing structures ordinarily found on more anterior segments. In the work that would eventually win the Nobel, Ed Lewis went on to discover a series of developmental mutations that disrupted the fly’s normal segmentation pattern, often causing anterior structures to shift toward the rear.

Mutational studies suggested that several of the genes responsible for these shifts in cell fate determination were linked together in the genome, forming what became known as the bithorax complex. The genetics also suggested that the Polycomb protein normally represses bithorax gene expression, keeping the genes off in body segments where their products don’t belong. This prevents structures such as sex combs or wings from forming in the wrong body segments.

Indeed, this is how the fly permanently shuts down these developmental genes. Poly-



Gene turnover. The methyl groups added to histone 3 of chromatin by the Polycomb group complex PRC2 attract PRC1, which then shuts down nearby gene activity.

developmental control genes that establish the fates of specific cells in the developing embryo. Often this suppression—which occurs once the developmental control genes have done their work—is permanent and heritable, passed down to all those cells’ daughters throughout the life of the organism. “How can you keep something off for the lifetime of the organism?” asks biochemist Robert

comes in to block transcription of the marked genes into messenger RNA, although there is controversy about how they actually do this.

Biologists are now finding that Polycomb group proteins affect other important developmental events besides cell fate determination. They are apparently needed to maintain the stem cells that form and replenish the body’s tissues. They also help inactivate one

comb can maintain gene repression for the life of the fly, says Jeffrey Simon of the University of Minnesota, Twin Cities. And the original Polycomb is not alone in this gene repressive activity. Over the years, fruit fly geneticists identified several more genes that can, when mutated, produce similar shifts in segmental structures, indicating that they, too, suppress bithorax and other gene activities.

Today, the *Polycomb* group of genes has some 15 members. The others were also discovered on the basis of their mutational effects on flies, and for the most part they are not structurally related to one another. They are widely distributed in nature, however. *Polycomb* group genes “are found in organisms from flies to humans,” Simon says. “Nearly every one is conserved.”

Uncovering the mechanism

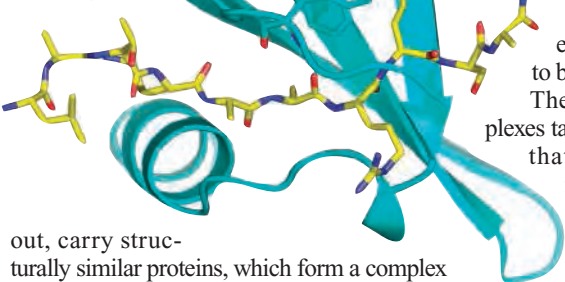
Although intriguing, the fruit fly mutation studies could not provide insights into how Polycomb group proteins shut down gene activity. Researchers needed to get their hands on the actual genes, but the first *Polycomb* group gene wasn't cloned until 1991 when Renato Paro, then a postdoc in David Hogness's lab at Stanford University School of Medicine in California, achieved the feat for *Polycomb* itself.

Analysis of the gene sequence provided the first clue to the Polycomb protein's modus operandi. The gene encodes a protein with a stretch of 37 amino acids that is similar to a known chromatin-binding domain in a protein called HP1 (for heterochromatin-associated protein 1). That suggested that Polycomb interferes with gene activity by attaching to chromatin in some fashion.

Shortly thereafter, researchers, including Simon, Paro, who is now at the University of Heidelberg, Germany, and Vincenzo Pirrotta, who recently moved from the University of Geneva, Switzerland, to Rutgers University in Piscataway, New Jersey, identified DNA sequences called Polycomb responsive elements (PREs). These are base sequences that are necessary for the repression of nearby genes by Polycomb group proteins. The assumption is that the sequences help attract the proteins to the right genes. Although uncer-

tainties remain, researchers have recently built a picture of how that happens.

In particular, they've shown that gene inactivation requires the cooperation of two complexes of the various Polycomb group proteins. The first, called PRC1 (for Polycomb repressive complex 1), was isolated from the fruit fly about 5 years ago by Kingston, Nicole Francis, who is also at Harvard, and their colleagues. PRC1 contains four core proteins—Polycomb itself plus PH (polyhomeotic), PSC (posterior sex combs), and dRING1—and binds to chromatin. Once there, it blocks the effects of a known gene-activating protein complex called SWI/SNF. Humans, it turns



out, carry structurally similar proteins, which form a complex with similar activity. PRC1 “seems to be the engine of [gene] repression,” Kingston says.

The identification of a second complex of Polycomb group proteins, PRC2, provided a major insight into how PRC1 knows which genes to target. In 2002, four groups, those of Kingston, who was working with Simon and Jürg Müller of the Max Planck Institute for Developmental Biology in Tübingen, Germany, Pirrotta, Danny Reinberg of the University of Medicine and Dentistry/Robert Wood Johnson Medical School in Piscataway, and Yi Zhang of the University of North Carolina, Chapel Hill, came across PRC2 more or less simultaneously.

The key observation about this complex was that one of its components, known as E(Z) for Enhancer of Zeste, has the ability to add methyl groups to the amino acid lysine 27,

which is located in the tail at the end of histone 3 of chromatin. Much evidence acquired over the past several years has shown that histone modifications play a major role in regulating



Developmental regulator. The Polycomb protein (cyan), a portion of which is shown here bound to a histone (yellow), helps ensure that structures like sex combs (*left, in the micrograph*) develop on correct fruit fly body segments.

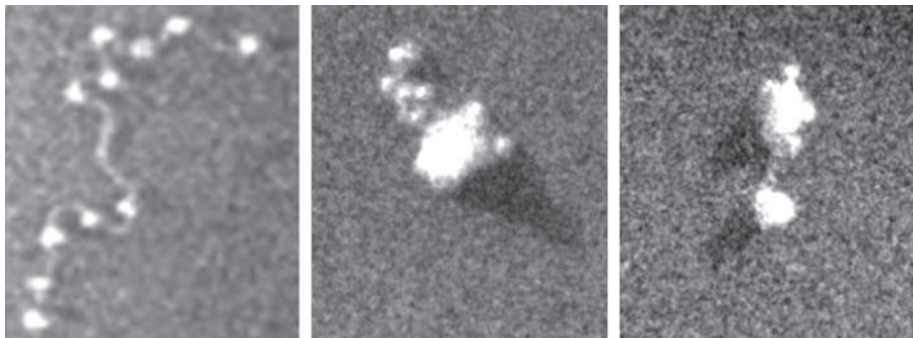
the activity of genes, turning them either on or off, depending on the modification. In PRC2's case, the methyl addition turns genes off, apparently by attracting PRC1 to the genes to be inactivated.

The researchers found that both complexes target the same chromosomal sites and that PRC2's methylating activity is needed for PRC1 binding. When PRC2 methylates histone 3, it's “like putting a little signpost in the chromatin that says ‘PRC1 bind here,’”

Simon explains. Although there is still some uncertainty about how PRC2 finds the right chromatin regions to tag, a team including Richard Jones of Southern Methodist University in Dallas, Texas, Judith Kassis of the National Institute of Child Health and Human Development in Bethesda, Maryland, and Zhang have identified proteins that interact both with PREs and with PRC complex proteins that might possibly be involved in such targeting.

Some uncertainties

Another outstanding issue for Polycomb researchers concerns how PRC1 inhibits gene activity. The simplest possibility is that it compacts the chromatin structure so that the transcribing machinery can't get access to the gene. There is some evidence for this. Isolated chromatin looks something like beads on a string; the beads are the so-called nucleosomes, consisting of DNA wound around a cluster of histone proteins, and the string is additional DNA that links the nucleosomes. Last year, Kingston, Francis, and Christopher Woodcock of the University of Massachusetts, Amherst, used electron microscopy to show that PRC1 compacts such nucleosome arrays, apparently causing the beads to clump together to the point at which they can no longer be distinguished. The researchers found that this compaction requires a segment of PSC, one of PRC1's core proteins that is needed for



Tied up in knots. When not condensed, chromatin exists in a “beads on a string” conformation (*left*). But when treated with PRC1, the beads clump together (*middle, right*).

gene repression, a result indicating that the two activities are linked.

But other researchers, such as Pirrotta, aren't so sure that PRC1 works simply by condensing the chromatin and thus blocking out the transcription machinery. Using a standard reporter gene assay for PRC1-mediated silencing, he and his colleagues recently showed that such silencing doesn't prevent binding by RNA polymerase, the enzyme that copies the DNA into messenger RNA. Instead, PRC1 apparently keeps the polymerase from transcribing the gene. "When we looked at the promoter [where the enzyme binds], RNA polymerase is there, but it can't get moving and open the DNA strands" to allow transcription, Pirrotta says. More work will be needed to clarify this issue, but Kingston, for one, suggests that both mechanisms, DNA compaction and inhibition of the transcription machinery, might conceivably come into play.

Although methyl addition to histone 3 by Polycomb group proteins can clearly tag genes for inactivation, the finding doesn't explain what makes the inactivation permanent. "The repressed state remains over many mitotic [cell] divisions. How is it maintained during DNA replication?" Paro asks. Recent results from his lab suggest a possibility.

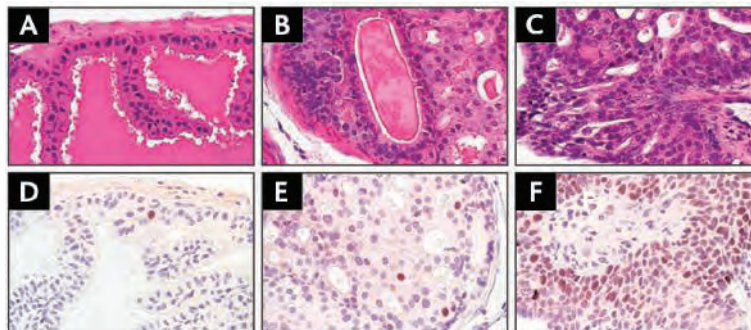
In work published online on 1 March in *Genes and Development*, the Heidelberg workers described evidence suggesting that Polycomb inactivation of PRE-associated genes occurs continuously unless something intervenes to prevent it. Thus, the silenced state could be maintained throughout the lifetime of the organism. But obviously, not all of these genes are shut down during development. Some remain "on" to produce the fly's normal segmental structures and perform other cellular functions. The Paro group has evidence that this active state is enabled by ongoing transcription of the PRE sequences, which somehow prevents Polycomb-mediated silencing, possibly because the transcription alters chromatin structure in such a way as to block Polycomb binding.

A broader view

Recent work suggests that the developmental significance of Polycomb group proteins goes far beyond their effects on bithorax gene expression. For example, the proteins contribute to normal development by helping inactivate one of the two X chromosomes carried by female cells. Two years ago, Zhang's group and independently, those of Neil Brockdorff of Hammersmith Hospital in London and Thomas Jenuwein of the Research Institute of

Molecular Pathology in Vienna, showed that such X inactivation depends on PRC2. Among other things, the researchers found that the complex binds to an X chromosome when inactivation begins and that PRC2-mediated methylation is needed to stabilize the chromatin structure of the inactive X.

The Polycomb group proteins also have roles beyond developmental regulation. By surveying the fruit fly genome for PRE sequences, Paro and his colleagues identified more than 150 genes throughout the genome that could be subject to Polycomb repression. Among these were various genes involved in controlling cell growth and division.



Cancer indicator? Micrograph A shows normal prostate epithelium, B shows a precancerous lesion, and C, full-fledged cancer. As the cancer progresses, EZH2 expression (purple in D, E, and F) increases.

Consistent with that, researchers have recently linked anomalies in Polycomb group gene expression with cancer development and progression. In particular, Arul Chinnaiyan of the University of Michigan Medical School in Ann Arbor, Mark Rubin of the Dana-Farber Cancer Institute in Boston, and their colleagues have looked at the expression of EZH2, the human equivalent of the fruit fly *E(z)* protein, in prostate and breast cancers. They found that the expression is much higher in cancers that have spread (metastasized) to other tissues than it is in localized tumors or normal tissue. Working with a mouse model of prostate cancer, Reinberg and his colleagues have confirmed that EZH2 production goes up as the cancers progress from localized to metastatic.

Increased EZH2 expression may in fact be a much-needed prognostic indicator for prostate cancer. Although many men develop small, localized prostate tumors as they age, most of these never progress and metastasize. "Most people die with [prostate cancer] rather than of it," Chinnaiyan says. But some of those localized tumors will metastasize, and currently it's impossible to identify the dangerous ones. This means that men may have to undergo therapy unnecessarily, and that can produce unpleasant side effects such as incontinence and impotence.

But in a small study of surgically removed human prostate cancers, published in the 10 October 2002 issue of *Nature*, the Chin-

naiyan team found that increased EZH2 expression in small, localized tumors was associated with a high risk of eventual disease spread. The overexpression "portends aggressiveness and metastasis," Chinnaiyan says. He and his colleagues are now organizing a larger clinical trial to confirm these preliminary findings. In addition, the protein may even provide a target for anticancer drugs. Chinnaiyan and colleagues have found that blocking production of the protein inhibits the proliferation of prostate cancer cells.

How EZH2 overproduction contributes to cancer development remains murky, but one possibility is that it disturbs normal gene control. Because Polycomb group proteins mainly repress genes, a flood of EZH2 may inhibit tumor-suppressor genes or genes that make proteins that keep cells anchored in place so that they can't migrate to new tissues as metastatic cells do.

Another clue comes from Reinberg and his colleagues. They found that EZH2 overproduction leads to formation of a Polycomb protein complex that differs in protein composition from PRC2. This could also lead to changed patterns of gene expression, he suggests.

Intriguingly, EZH2 overexpression and formation of the PRC variant occurs in undifferentiated cells as well as in cancer cells. This is consistent with the views of some researchers that cancer cells behave as if they have regressed to a more primitive developmental state. It is also consistent with recent findings by Jenuwein, Azim Surani of the Wellcome/CRC Institute of Cancer and Developmental Biology in Cambridge, U.K., and others suggesting that histone methylation mediated by EZH2 helps maintain stem cells in their pluripotent developmental state.

The Polycomb group proteins are clearly turning out to be highly versatile players in a wide range of cellular activities. And still more revelations may be in store. Within the past year, researchers including Brockdorff and Zhang have reported that some Polycomb group proteins can add the small protein ubiquitin to histone H2A. Originally discovered as a tag that marks proteins for destruction, ubiquitin has since been shown to have many other roles in the cell, including regulation of gene expression and protein migrations (*Science*, 13 September 2002, p. 1792).

The Polycomb-mediated histone ubiquitination is involved in gene silencing, but Zhang says its exact role isn't yet known. One thing is clear, however. At 60 years of age, the Polycomb group proteins are still showing plenty of life. **-JEAN MARX**

Kansas Gears Up for Another Battle Over Teaching Evolution

Scientists plan to avoid hearings by the Kansas Board of Education on intelligent design and evolution. But they hope that economic arguments will carry the day

LAWRENCE, KANSAS—This month, after voters overwhelmingly approved a constitutional amendment making Kansas the 18th state to ban gay marriages, Reverend Jerry Johnston announced that his next targets were evolution, gambling, and abortion. Over the next 3 weeks, the pastor of the rapidly growing First Family Church in Overland Park in northeast Kansas delivered sermons attacking Darwin's theory and lauding intelligent design (ID), the idea that a higher intelligence played a role in creating life on Earth. "Getting intelligent design into school curricula is the worthiest cause of our time and the key to reversing the country's moral decline," says Johnston. "The evangelical and ID communities must work together to make that happen."

That prospect sends chills down the spines of most Kansas scientists and educators. They are already dreading the publicity that is likely to accompany 6 days of hearings next month by the Kansas State Board of Education—a majority of whose members are ID supporters—to kick off the process of revising state science standards for all Kansas students. The scientific community plans to boycott the hearings, calling them a "kangaroo court," but it isn't ignoring Johnston and his followers. Last week more than 100 people opposed to making ID part of the science curriculum held a meeting in a liberal church here to test a new rallying cry: A high-quality science education means more jobs and a stronger economy. By attracting business, civic, and religious leaders, supporters hope to erode ID's traditional base and stave off changes that they believe will make Kansas an undesirable location for high-tech companies, academics, and other knowledge-based workers.

"We need to turn K-12 education in Kansas into a powerhouse producer of science-literate students," says biologist Steve Case of the University of Kansas, Lawrence, chair of the board's 26-member science standards writing committee and a speaker at the meeting. "Teaching intelligent design would do the opposite."

The recent events are part of a seesaw battle over the science curriculum in Kansas. State standards were revised in 1999 to make room for ID. But those changes were reversed 2 years later, after voters booted out some of the more conservative members on the

10-person board. Last November, however, evangelicals and ID proponents led by John Calvert, a managing director of the Intelligent Design Network in Shawnee Mission, Kansas, helped propel conservatives back into the majority, setting off a new push to revise the standards.

Within a month, a minority within the state standards writing committee proposed



From the pulpit. Steve Case and other Kansas scientists hope to make religious leaders allies in the debate over intelligent design.

changing the definition of science to include explanations other than "natural" and to insert the proposition that evolution was "a theory, not a fact." "It was a complete subversion of the process," says Case, who describes the ID backers as having shown the "tenacity of pit bulls." Although Case told the state board that the proposed changes had been soundly defeated within the committee as part of its deliberations, the board decided to hold hearings on the issue.

"We feel that this is a legitimate scientific controversy that needs to be laid out on the table," says Kathy Martin, one of the three members on the panel that will preside over the 5-7 and 12-14 May hearings. She

says the proceedings will likely lead to "certain revisions" in the science standards.

Sue Gamble, a board member who opposes the inclusion of ID in science teaching, admits that her side let down its guard after the state standards were revised in 2001. At the same time, she says, "evangelical megachurches galvanized their parishioners into a formidable voting bloc that views evolution as part of a whole amalgam of issues that include gay marriages and abortion."

The 21 April meeting here is part of a belated attempt to catch up, say evolution supporters. The site—Plymouth Congregational Church, one of the state's oldest and most liberal churches—was intended to send a message that the teaching of evolution is compatible with religious doctrine. "Some people have

the mistaken notion that science and faith are at loggerheads. But there are vibrant Christian communities here that understand that the Bible is not a scientific text," says Plymouth's pastor, Peter Luckey.

John Burch, a local investor who organized the meeting with help from the nonprofit Kansas Citizens for Science, warned that introducing ID in school curricula would undermine a state-backed plan to invest \$500 million over the next 10 years to boost Kansas's bioscience industry. "Most industries today want workers with analytical skills," says microbiologist Charles Decedue, executive director of the Higuchi Biosciences Center at the University of Kansas, which is dedicated to the development and transfer of bioscience technologies. "ID does not foster analytical thinking because its arguments are faith-based."

Don Covington, a vice president of the Intelligent Design Network, is unimpressed by the economic argument. "Corporate executives don't discuss Darwinism while discussing business projects," says Covington, who was one of a half-dozen ID supporters in the audience. As for ID instruction keeping families away from the state, he says that when "kids find out that they are going to learn the truth, they might be excited to come here."

Burch hopes to win more support from industry by meeting with researchers and executives at local bioscience companies. And he plans to keep the message simple. "Evolution versus intelligent design is not a scientific issue," he says. "It's a workforce issue."

—YUDHIJIT BHATTACHARJEE

RANDOM SAMPLES

Edited by Constance Holden

Monkeys Strike Gold

Last year, Robert Wallace, a researcher with the Wildlife Conservation Society (WCS), discovered a new species of tiny monkey in the Bolivian jungle. Now he's raised money to sustain the creature's habitat, Madidi National Park, by auctioning off the rights to name it. Earlier this month, WCS announced the results of the online auction: An Internet casino called GoldenPalace.com won with a bid of \$650,000. The monkey will henceforth be known as *Callicebus aureipalatii* (Latin for golden palace). Little is known about the diminutive fructivores except that—as depicted here—they like to hang onto each other and holler in the morning.



New Cambridge Center Emerges

"Emergence"—the idea that things are more than the sum of their parts—is "one of the most compelling new concepts in science," according to the John Templeton Foundation in Conshohocken, Pennsylvania. It's used to explain everything from the coalescence of dust into stars to the rise of intelligent organisms.

So the foundation is supporting a new group at Cambridge University, the Cambridge Templeton Consortium, which starting this summer plans to hand out \$3 million in grants for research on the emergence of complex systems in three areas: biochemistry, evolution, and cognition.

Inspired by the idea that the universe

would have been a nonstarter if fundamental physical constants were slightly different, the consortium wants to look for similar fine-tuning in biology. "I am convinced that there are deep structures in biology, and evolution navigates over them," says paleontologist Simon Conway Morris, the consortium's director. If such structures exist, he holds, humans might still emerge if evolution had to start over again on Earth, and life on other planets could be much like ours.

Molecular biologist Steven Benner of the University of Florida, Gainesville, applauds the initiative. "There is a strong need in the biomolecular sciences" to address questions such as "Why is life the way that it is?" he says. Others are skeptical. "I don't think that a scientific-theological-philosophical mélange is going to make a significant contribution" to scientific knowledge, says

paleobiologist Doug Erwin of the Smithsonian Institution in Washington, D.C. It's worth a shot, though, argues Conway Morris: "This is very much an experiment."

Sex and Science (cont.)

The debate goes on. A panel of scientists convened by the New York Academy of Sciences met in New York on 14 April to kick around some of the dust raised by Harvard President Lawrence Summers last January when he suggested that biological sex differences might have something to do with why there are fewer women than men in science.

"Way out on the end of the bell curve is where this controversy lies," said Richard Haier, a psychologist at the University of California, Irvine, pointing out that males dominate at the extreme reaches of math achievement, vastly outnumbering females among those who score above 700 on the math SAT. Joshua Aronson, a psychologist at New York University, countered that stereotypes dramatically affect test performance. Studies have shown that women do better on math tests in an all-women test group, he said: "Even one man in the room made the women's scores drop."

Diane Halpern, a psychologist at Claremont McKenna College in Claremont, California, observed that differences aren't necessarily deficiencies. "Maybe we should be asking what is holding men back? They get only 32% of the Ph.D.s in psychology." Sociologist Linda Gottfredson of the University of Delaware, Newark, cited research showing women prefer fields that deal with people rather than things. "Why do we want equal proportions of men and women in each profession?" she asked. Nancy Hopkins, the Massachusetts Institute of Technology biologist who helped raise the dust in the first place, remained unmoved, saying the numbers would be different "if the door were truly open."

Champion Racer Cloned

This month, a French-Italian collaboration announced the successful birth of a foal cloned from a gelding. Now 2 months old, the foal was produced by the French genetic engineering company Cryozotech and the Italian reproductive technology lab LTR-CIZ.

The lab's team, headed by Cesare Galli, has improved on techniques it used 2 years ago to produce the first horse clone, a mare. From 200 nuclear transfers using skin cells from Pieraz, a retired thoroughbred Arabian endurance champion, the researchers got 34 embryos and three pregnancies, one of them successful.

Galli has predicted that cloning will revolutionize the horse-racing industry. But at present, the thoroughbred racing community doesn't even permit artificial insemination, much less cloning. Paul Struthers of Britain's Jockey Club says racers have a very restricted gene pool and "there would be very serious implications for the long-term welfare of the thoroughbred were the gene pool to be reduced further" by breeders all going after the progeny of superachievers. But cloning could have a future with horses intended for show jumping, dressage, or endurance racing—events with fewer breeding restrictions. Over 90% of dressage stallions are gelded to make them more manageable, says Nicolas Robin of Cryozotech: "So imagine how many gene lines are lost." But no longer. The company is preserving cells from some 30 prize stallions and plans to market semen from their clones.



Edited by Yudhijit Bhattacharjee

AWARDS

Tech wonders. Elwood "Woody" Norris says there's a trick to being a good inventor: Find something commonplace in one area of science and extend it to another sector. Last week that trick paid off big, with Norris winning the \$500,000 Lemelson–Massachusetts Institute of Technology (MIT) prize from the Lemelson Foundation.

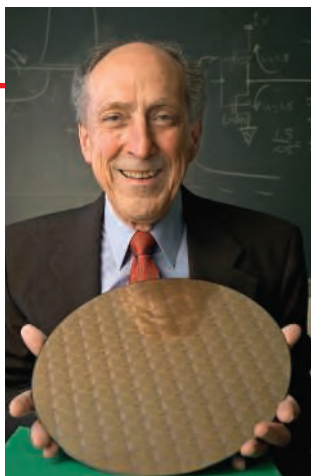
Norris (below), 67, of San Diego, California, was honored for his numerous inventions,



including a system that focuses sound waves akin to the way a common flashlight focuses light. The invention, called Hyper-Sonic Sound, is being used to

target advertising to individual customers in a crowd without disturbing their neighbors.

Norris has also worked on making helicopters cheaper and easier to operate. His one-person helicopter, dubbed the AirScooter, is expected to start selling commercially this fall for \$50,000 each.



Last week the foundation also gave its \$100,000 lifetime achievement award to IBM fellow Robert Dennard (above), inventor of dynamic random access memory, the "working memory" for most computers. That invention—along with the integrated circuit, the metal oxide field effect transistor, and magnetic hard disk—is one of the critical components of modern computing, says MIT electrical engineer Dimitri Antoniadis.

IN THE NEWS

Reaching out. One silver lining in the HIV/AIDS pandemic, says public health guru Fitzhugh Mullan, may be an increased U.S. awareness of its responsibility to address global health issues. Last week, a panel from the National Academies' Institute of Medicine

PIONEERS

Not alone. As one of only two women to lead a university in the Arab world, epidemiological psychiatrist Rafia Ghubash belongs to an exclusive club. But the 49-year-old president of the Arabian Gulf University (AGU) in Bahrain is working hard to lose that status.

This month Ghubash launches a network to help more women scientists attain leadership positions and to attract more women into science. The initial goal will be "to simply make them aware of how many they are and provide role models," says Ghubash. That awareness, she hopes, will counter pressure on women with undergraduate science degrees to become "teachers, nurses, or to drop out entirely to raise families."

Only a few hundred women from the 22 Arab countries have signed up for the network so far, but Ghubash expects "at least a million" after its official launch at a meeting of women scientists at AGU on 15 May. A Web site based at AGU will hold forums on issues such as gender bias and feature a scientific newsletter. In a few months, women will be able to post their CVs online and hunt for scholarships and jobs.



(IOM) that he chairs proposed a way to institutionalize that commitment through a U.S.-funded program to help 15 impoverished countries cope with the disease.

"I've waited 25 years for an opportunity like this," says Mullan, 62, a former assistant

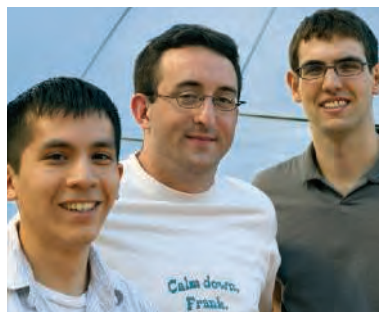
U.S. surgeon general now multitasking as a policy analyst, author, and professor of pediatrics and public health at George Washington University in Washington, D.C. The IOM report, *Healers Abroad*, calls for a U.S. Global Health Service staffed by a 150-member corps of experienced medical professionals and supplemented with as many as 2000 fellows—recent graduates working off student loans as well as those interrupting their careers—who would be deployed throughout the countries targeted by the President's Emergency Plan for AIDS Relief (PEPFAR). Mullan calls the \$140-million-a-year price tag a "drop in the bucket" for the multibillion dollar PEPFAR and says he'd "be very disappointed" if the U.S. State Department, which requested the report, doesn't adopt most of the panel's recommendations within a year.

SNAFUS

Junk science. Fed up with spam from a computer science conference soliciting papers, Jeremy Stribling (center) and two other graduate students, Dan Aguayo (left) and Max Krohn (right), at the Massachusetts Institute of Technology hit pay dirt by successfully submitting computer-generated gibberish. The technique uses context-free grammar, which rearranges sentences in a way that is grammatically correct but makes no sense.

"We suspected that their standards were low," Stribling says of the World Multiconference on Systemics, Cybernetics and Informatics, which accepted their paper, "Rooter: A Methodology for the Typical Unification of Access Points and Redundancy." Following media coverage of the prank, the organizers of the conference, to be held this summer in Orlando, Florida, rejected the paper and refunded the authors' registration fee. General chair Nagib Callaos said on the group's Web site that "the acceptance of a small percentage of nonreviewed papers does not significantly decrease the quality level of a conference."

Stribling's group plans to continue their research at the conference. "We plan to give a randomly generated talk," he says. "Even we won't know the slides when we get there."



Got any tips for this page? E-mail people@aaas.org

CREDITS (TOP TO BOTTOM): PHOTO BY ALAN ORLING; COURTESY OF IBM; SOURCE: ARABIAN GULF UNIVERSITY; SOURCE: MIT; FRANK DABEK

A Cry for Help from Kansas

DEAR LADIES AND GENTLEMEN OF THE American Science Community: As I write this from my troubled home state of Kansas, the State Board of Education is debating once again whether to de-emphasize the theory of evolution. Those in the majority on the Board have stated that they are considering changing the very definition of science to allow for science classes to discuss the merits of intelligent design. Kansas Attorney General Phill Kline has said that he would defend the Board's actions.

Kansas may be a far-off, conservative state for most of you, largely unimportant in the affairs of your professional lives, but this is a strong indicator that the very foundation of science in the United States is at risk.

“What a shame it would be if unqualified politicians succeed in undoing centuries of scientific progress in both the public's perception of science and its continuing advancement.”

—REYNOLDS

Other states have had this problem recently, including Georgia, Alabama, and Ohio, and this could happen to your state. What a shame it would be if unqualified politicians succeed in undoing centuries of scientific progress in both the public's perception of science and its continuing advancement. This is a wake-up call, ladies and gentlemen. Please don't let this happen in your state.

You may write to the Board by visiting their Web site at www.ksbe.state.ks.us/Welcomed.html.

ERIC REYNOLDS

Overland Park, KS, USA.

What Causes Lesions in Sperm Whale Bones?

THE DESCRIPTION OF “EXTENSIVE PROLIFERATION and remodeling of cartilage and woven bone” in sperm whales in the intriguing Brevia by M. J. Moore and G. A. Early (“Cumulative sperm whale bone damage and the bends,” 24 Dec. 2004, p. 2215) is not compatible with their diagnosis of avascular

necrosis (osteonecrosis). Avascular necrosis, attributed to bends, has been clearly documented in extinct whales (1). Absence of reactive new bone formation around areas of dead bone is visualized as intraosseous clefts or as articular subsidence, producing a depressed zone (2–5). Moore and Early's description is compatible with another known cetacean disease, spondyloarthropathy (5, 6). This disorder is especially common in the cetacean *Lagenorhynchus*, and zygapophyseal joint erosions of spondyloarthropathy have also been found in a blue whale (7).

The reported frequency of sperm whale “avascular necrosis” is compatible with the 100% reported in certain genera of extinct mosasaurs (2), but substantially greater than the frequency of spondyloarthropathy found in other mammals (5). While Moore and Early's

diagnosis of avascular necrosis cannot be substantiated, the pathology frequency they suggest is also outside anticipated ranges (5 to 50%) for spondyloarthropathy in mammals (5, 8). Examination of their fig. 1 clearly explains the apparent variation. The figure illustrates bone character underlying normal articular surfaces and the subchondral erosions of spondyloarthropathy (5, 9). The undulating surface with visible pores (in most of the sections) is characteristic of normal subchondral bone. Also illustrated is erosive damage with reactive new bone formation, characteristic of spondyloarthropathy (4, 5, 8).

BRUCE M. ROTHSCHILD

Arthritis Center of Northeast Ohio, Northeastern Ohio Universities College of Medicine, 5500 Market Street, Youngstown, OH 44512, USA.

References

1. B. M. Rothschild, A. Saunders, *J. Vertebr. Paleontol.* **17**, 72A (1997).
2. B. M. Rothschild, L. D. Martin, *Science* **236**, 75 (1987).
3. B. M. Rothschild, G. Storrs, *J. Vertebr. Paleontol.* **23**, 324 (2003).
4. D. Resnick, *Diagnosis of Bone and Joint Disorders* (Saunders, Philadelphia, PA, 2002).
5. B. M. Rothschild, L. D. Martin, *Skeletal Impact of Disease* (New Mexico Museum of Natural History Press, Albuquerque, NM, 2005).
6. E. O. Kompanje, *Z. Säugetierkd.* **58**, 316 (1993).
7. B. M. Rothschild, V. Naples, *Ann. Rheumat. Dis.*, in press.
8. B. M. Rothschild, C. Rothschild, *Paleontol. Soc. Publ.* **8**, 330 (1996).
9. A. Boyde, E. C. Firth, *J. Anat.* **205**, 491 (2004).

IN THEIR BREVIA “CUMULATIVE SPERM WHALE bone damage and the bends” (24 Dec. 2004, p. 2215), M. J. Moore and G. A. Early interpret ontogenetically progressive, chronic osteonecrosis in specimens collected over 111 years as being caused by nitrogen emboli—

with variations from or exogenous disruptions (anthropogenic sonar) to diving patterns possibly causing acute embolic disease.

However, their sample is from the post-1860s mechanized whaling era, with exposure to effects of both luse-jag—stalking to surprise prey—and prøysser-jag—persistent direct chasing to run down prey. An alternative hypothesis is that all the observed effects are related to human impacts and are not

Image not available for online use.

Socializing sperm whales

biologically normal. To test both hypotheses, samples from open boat whaling (1710s to 1860s) that exploited the diving cycle of large bulls (1), historic pre-exploitation strandings, and prehistoric fossils must be studied. Anthropogenic impacts also might be causal in nonlethal implosions of mysticete auditory bullae due to dysbaric conditions (2).

EDWARD D. MITCHELL

Natural History Museum of Los Angeles County, 900 Exposition Boulevard, Los Angeles, CA 90007, USA. E-mail: edmnhm@ix.netcom.com

References

1. E. Mitchell, *Rep. Int. Whal. Comm. Spec. Iss.* **5**, 63 (1983).
2. F. C. Fraser, P. E. Purves, *Scott. Nat.* **65**, 154 (1953).

Response

IN RESPONSE TO ROTHSCHILD'S SUGGESTION that the sperm whale bone lesions we described as osteonecrotic are pathognomonic for spondyloarthropathy, we offer the following observations: The extensive proliferation and remodeling of cartilage and woven bone that we described in sperm whales is a central aspect of experimentally induced dysbaric osteonecrosis in sheep (1). However, in recent correspondence with Rothschild, he kindly shared with us two pertinent papers in press (2, 3) that describe erosive subchondral lesions diagnosed as spondyloarthropathy somewhat comparable to some of the lesions we described in sperm whale bones. Our diagnosis of osteonecrosis can, of course, be questioned; however, the selective benefits of such a high frequency of spondyloarthropathy are obscure. In con-

trast, any dysbaric cost could well be outweighed by the trophic advantage of access to deep-sea squid and fish.

We remain puzzled by the large sub-spherical bubble-like cavities, in rib and chevron bones, illustrated in our paper. These lesions do not appear to be comparable to published descriptions of spondyloarthropathy, nor do the bizarre nasal bone changes also illustrated.

We therefore suggest that the diagnosis of osteonecrosis needs to be confirmed with further histological and radiographical study of nonautolyzed material from future sperm whale mortalities.

Mitchell suggests that the progressive osteonecrotic lesions observed in sperm whales over 111 years are best interpreted as resulting from lesions induced by whaler harassment and are thus human impacts. We agree that further studies of earlier whaling, stranding, and fossil cases are warranted to test the alternative hypothesis. We are aware of two pre-1860s stranded specimens: one described by Melville (4) stranded in Tunstall, Yorkshire, UK, in 1825, and another that was beached and killed in 1661 at Easington, County Durham, UK (5). In the former case, weathering precludes diagnosis of bone condition. Photographs of the latter show no obvious lesions. Nevertheless, we will continue to pursue Mitchell's concerns and report further as warranted.

MICHAEL J. MOORE AND GREG A. EARLY

Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA.

References

1. T. F. Lin *et al.*, *Undersea Hyperbaric Med.* **23**, 39 (1996).
2. B. Rothschild, F. Ruhli, *Am. J. Primatol.*, in press.
3. B. Rothschild, F. Ruhli, *Am. J. Primatol.*, in press.
4. H. Melville, *Moby Dick* (Harper & Brothers, New York, 1851), chap. 102.
5. J. Drury, *Durham County Local Hist. Soc. Bull.* **39** (15 Dec. 1980).

Ethics of Tobacco Company Funding

AS ONE OF THE RESEARCHERS INTERVIEWED BY David Grimm for his article on tobacco company funding of research ("Is tobacco research turning over a new leaf?", *News Focus*, 7 Jan., p. 36), I felt that it was an informative and well-written piece. I took some exception, however, to the last sentence, which construed not accepting funding from tobacco companies as taking "the moral high road." Those of us who accept tobacco company funding with the aim of saving lives through our research are just as much on the "moral high road" as those who oppose tobacco funding. Our research, using tobacco company funding, is dedicated to the development of improved smoking cessation treatments, which will help save lives (1–3).

Federal sources of funding for research on tobacco dependence are all too scarce. Even those who vehemently oppose taking money directly from tobacco companies take money indirectly from tobacco companies, such as that from the Master Settlement Agreement. However, sadly, far too little of the funding from the Master Settlement Agreement is going into the intended uses of supporting

“

Awarding grants directly from tobacco companies to university-based research groups provides better assurance that the money will be used to combat tobacco addiction and related problems.”

—ROSE

research and treatment of tobacco addiction. Awarding grants directly from tobacco companies to university-based research groups provides better assurance that the money will be used to combat tobacco addiction and related problems. Understandably, safeguards must be set in place to prevent tobacco companies from controlling the direction of the research or censoring the publication of results. We have instituted such safeguards in our research program.

Current long-term success rates with smoking cessation treatments in real-world settings are dismally low, often less than 20% (4). Tobacco company funding can be instrumental in the development, evaluation, and dissemination of more effective treatments.

JED E. ROSE

Director, Center for Nicotine and Smoking Cessation Research, Duke University Medical Center, 2200 West Main Street, Suite B-150, Durham, NC 27705, USA. E-mail: rose0003@mc.duke.edu

References

1. J. E. Rose, E. D. Levin, F. M. Behm, C. Adivi, *Clin. Pharmacol. Ther.* **30**, 323 (1990).
2. J. E. Rose, F. M. Behm, E. C. Westman, *Exp. Clin. Psychopharmacol.* **6**, 1 (1998).
3. E. C. Westman, F. M. Behm, J. E. Rose, *Chest* **107**, 1358 (1995).
4. S.-H. Zhu, T. Melcer, J. Sun, B. Rosbrook, J. P. Pierce, *Am. J. Prev. Med.* **18**, 305 (2000).

Merits of a New Drug Trial for ALS?

AS NEUROSCIENTISTS WORKING ON AMYOTROPHIC lateral sclerosis (ALS), we read with interest the Perspective "Treating neurodegenerative diseases with antibiotics" by T. M. Miller and D. W. Cleveland (21 Jan., p. 361) and the article on which it comments, by J. D. Rothstein *et al.* on the use of β -lactam anti-

otics to treat ALS (1). We are impressed by the screening of 1040 U.S. Food and Drug Administration–approved compounds for increased glutamate transport, which began in the spring 2002 and led to initiation of a clinical trial in ALS with ceftriaxone in spring 2005. We are concerned that the previous results of clinical research on ceftriaxone have been overlooked. In fact, a Medline

search identified eight negative trials with ceftriaxone in ALS patients, published between 1992 and 1996 (2–9).

The fact that ceftriaxone increased glutamate uptake in rat spinal cord synaptosomes has already been published (2), and this prompted us to verify the efficacy of the drug in 108 ALS patients, with disappointing results (2).

Thus, it seems that, although for basic science the new molecular biology approaches have provided elegant and specific contributions to explain the mechanism by which ceftriaxone increases glutamate uptake, from a clinical perspective, scientific research into ALS has actually not improved over the last 10 years. The emphasis on this drug as a possible solution to "one of the big challenges of this century" is not appropriate, considering the previous negative results.

ETTORE BEGHI,* CATERINA BENDOTTI,

TIZIANA MENNINI

Istituto di Ricerche Farmacologiche "Mario Negri," Via Eritrea, 62, 20157 Milan, Italy.

*To whom correspondence should be addressed.

E-mail: beghi@marionegri.it

References

1. J. D. Rothstein *et al.*, *Nature* **433**, 73 (2005).
2. The Italian ALS Study Group, *Eur. J. Neurol.* **3**, 295 (1996).
3. F. H. Norris, *Arch. Neurol.* **51**, 447 (1994).
4. J. R. Gil Llano, I. Casado Naranjo, *Neurologia* **9**, 205 (1994).
5. P. Couratier, J. M. Vallat, L. Merle, P. M. Preux, J. Hugon, *Therapie* **49**, 146 (1994).
6. V. Carelli, R. Liguori, C. Cordivari, G. Bianchedi, P. Montagna, *Ital. J. Neurol. Sci.* **15**, 66 (1994).
7. F. J. Carod Arta, I. Perez Lopez-Fraile, M. Gracia Naya, J. A. Giron Mombiola, *Neurologia* **9**, 29 (1994).
8. W. Robberecht, *Lancet* **340**, 1096 (1992).
9. L. G. Smith, *Lancet* **340**, 379 (1992).

Response

OUR PERSPECTIVE HIGHLIGHTED THE REMARKABLE discovery that a well-known penicillin derivative, ceftriaxone, was not only effective in fighting infection within the nervous system, but also induced (at the transcriptional level) the synthesis of the major glutamate transporter within the spinal cord (1). This transporter, which serves to quickly dampen chemical signaling from one neuron to another and thereby to limit repetitive, excitotoxic neuronal firing, is known to be lost in ALS, a disease that

causes severe paralysis from a progressive loss of motor neurons over a typical 2- to 5-year course. Beghi *et al.* correctly note that previous efforts (2–8) using short-term administration of ceftriaxone (between 1 and 8 weeks) in a small number of patients were conducted in the early 1990s after a case report claiming a remarkable benefit in a single patient (9). None of these short-term efforts were double-blinded, nor were placebo controls used, important trial design features that would be necessary for a persuasive outcome. In the most comprehensive of these (conducted by Beghi, Mennini, and others), one-third of the 21 patients treated with ceftriaxone for the longest period (5 to 8 weeks) were claimed to have shown improvement (8). These prior efforts do not dampen enthusiasm for the discovery of induction of glutamate transport by a drug already known to be safe and to penetrate the blood-brain barrier. A long-term, double-blinded, and placebo-controlled trial in ALS is just what the evidence warrants.

TIM MILLER AND DON CLEVELAND

Ludwig Institute for Cancer Research and the Department of Medicine and Neurosciences, University of California, San Diego, La Jolla, CA 92093, USA.

References

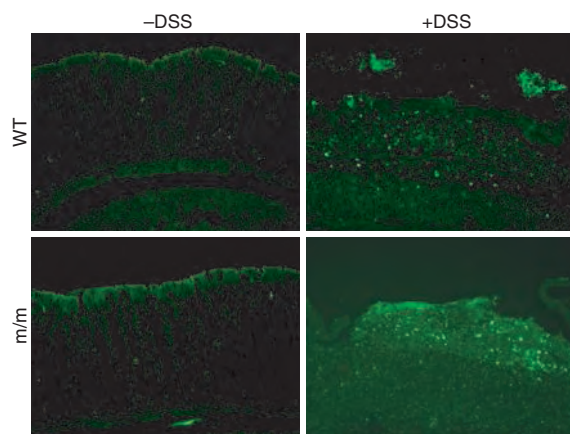
1. J. D. Rothstein *et al.*, *Nature* **433**, 73 (2005).
2. F. H. Norris, *Arch. Neurol.* **51**, 447 (1994).
3. J. R. Gil Llano, I. Casado Naranjo, *Neurologia* **9**, 205 (1994).
4. P. Couratier, J. M. Vallat, L. Merle, P. M. Preux, J. Hugon, *Therapie* **49**, 146 (1994).
5. V. Carelli, R. Liguori, C. Cordivari, G. Bianchedi, P. Montagna, *Ital. J. Neurol. Sci.* **15**, 66 (1994).
6. F. J. Carod Artal, I. Perez Lopez-Fraile, M. Gracia Naya, J. A. Giron Mombiola, *Neurologia* **9**, 29 (1994).
7. W. Robberecht, *Lancet* **340**, 1096 (1992).
8. The Italian ALS Study Group, *Eur. J. Neurol.* **3**, 295 (1996).
9. L. G. Smith, *Lancet* **339**, 1417 (1992).

CORRECTIONS AND CLARIFICATIONS

News Focus: "Mounting evidence indicts fine-particle pollution" by J. Kaiser (25 Mar., p. 1858). The article should not have described the lungs of asthmatics in a clinical study as "damaged." Lung function of these mildly asthmatic subjects was not affected by exposure to ultrafine particles. Asthmatics had 6 times more particles in their lungs than healthy subjects only when exercising asthmatics were compared to resting healthy subjects. Finally, many, but not all, short-term epidemiologic studies have linked sulfates and health effects. A study cited in the sidebar "How dirty air hurts the heart" appeared in the April/May issue of *Inhalation Toxicology*.

News of the Week: "Play and exercise protect mouse brain from amyloid buildup" by J. Marx (11 Mar., p. 1547). In the story, Gary Arendash was mistakenly identified as David Arendash. Also, Orly Lazarov's name was misspelled in the story and the image credit.

Reports: "Nod2 mutation in Crohn's disease potentiates NF- κ B activity and IL-1 β processing" by S.



Maeda *et al.* (4 Feb., p. 734). It was brought to the authors' attention that the two control panels in Fig. 4A look alike. After examining the issue, it was realized that during preparation of the figure, one of the control panels was mistakenly inserted twice, and the other control was omitted. The correct figure is shown here. The authors apologize for any possible confusion and inconvenience that may have been created. The results and conclusions remain as before, i.e., increased macrophage apoptosis in the lamina propria of *Nod2*^{2939ic} (m/m) mice after DSS treatment.

Policy Forum: "Do the largest protected areas conserve whales or whalers?" by L. R. Gerber *et al.* (28 Jan., p. 525). On page 525, in the second column, last paragraph, the heading should read, "Lack of an adaptive design." In the third column, second paragraph, the first sentence should read, "Current debate has polarized IWC members into those who advocate widespread sanctuary use and those who believe that they are redundant under the RMP/RMS [see supporting online material (SOM)]." In the same paragraph, the fourth sentence should read, "We assumed typical life history parameters for baleen whales (i.e., we used demographic parameters for the gray whale *Eschrichtius robustus*) (13, 14)." On page 526, first column, first full paragraph, the first sentence should read, "Although this result was robust to small (± 0.02) changes in parameter values, our model includes dispersal as diffusion, rather than as explicit migration, and does not consider density dependence, demographic stochasticity, or environmental noise." The second sentence in that paragraph should begin "However, given the uncertainties..." The third sentence should read, "Our results are consistent with results from previous work..." In the fourth to last line of this column, the citation should be (16) instead of (14). In the second column, first paragraph, the fifth sentence should

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 6 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

read, "A starting point would be the establishment of IWC sanctuaries conforming to more ecologically based designation." In the second paragraph, lines 5 through 7 should read, "populations of whales during certain time periods (e.g., in breeding grounds and/or feeding areas)..." In the third column, the first line should read "(SOM)" rather than "(22)." The second sentence in the third column should read, "Nevertheless, the adherence to a quota system would enhance whale conservation by restricting the times and areas of whale harvesting, and by limiting the total catch." In the References and Notes, in (2), the first line should read, "In accordance with the IWC, 'whaling' refers to the..." and the citation in the last line should be (22). In (9), the third author's initials appear twice (K.D.). Note (22) is now omitted, as the SOM is mentioned in the text, and (23) and (24) should become (22) and (23). The new note (23) [formerly (24)] should read, "We thank D. P. DeMaster and two anonymous reviewers for helpful comments, and the chair (A. Zerbini)..."

Perspectives: "The maser at 50" by R. L. Walsworth (8 Oct. 2004, p. 236). The figure legend is incorrect. The photo shows the second maser, not the first.

Perspectives: "NAD to the rescue" by A. Bedalov and J. A. Simon (13 Aug. 2004, p. 954). In reference 7, the first author's name was misspelled. It should be A. Sajadi.

TECHNICAL COMMENT ABSTRACTS

COMMENT ON "Ecosystem Properties and Forest Decline in Contrasting Long-Term Chronosequences"

Kanehiro Kitayama

Wardle *et al.* (Reports, 23 July 2004, p. 509) demonstrated that forest decline following a transient peak biomass is a common forest ecosystem dynamic caused by increased soil-phosphorus limitation over time. However, the decline they observed is attributable to the lack of phosphorus use-efficient species, and is confined to regions of low tree species diversity.

Full text at

www.sciencemag.org/cgi/content/full/308/5722/633b

RESPONSE TO COMMENT ON "Ecosystem Properties and Forest Decline in Contrasting Long-Term Chronosequences"

D. A. Wardle, L. R. Walker, R. D. Bardgett

Kitayama correctly recognizes that our study did not include hyperdiverse tropical forests. However, the data set he uses to test our ideas for tropical forests is not relevant to the spatial scale that we considered, and the mechanism that he proposes for these forests is not supported by current ecological theory.

Full text at

www.sciencemag.org/cgi/content/full/308/5722/633c

Comment on "Ecosystem Properties and Forest Decline in Contrasting Long-Term Chronosequences"

Wardle *et al.* (1) studied six long-term chronosequences in Australia, Sweden, Alaska, Hawaii, and New Zealand and found that in the absence of major ecosystem disturbance, a transient peak in forest biomass is commonly followed by a forest decline phase. They ascribed the decline of forest biomass to the decline of soil phosphorus (P) availability. With increasing substrate age, nitrogen to phosphorus (N:P) ratios of fresh litter and humus increased in the majority of chronosequences. The authors concluded that similar phases of forest decline occurred widely, from tropical to temperate to boreal forest ecosystems. I argue that their conclusion is premature and that their forest dynamics model does not take the function of species diversity into consideration.

Wardle *et al.* (1) report that forest decline is associated with increased P limitation relative to N and a reduced release of P from decomposing litter. Geochemical changes in soil phosphorus fractionation during long-term soil development have been well established (2). Subsequent biogeochemical studies using the Hawaiian Islands as a model system have proven the shift of N limitation in net primary productivity in the

developing phase to P limitation in the retrogressive phase (3–5). However, the decline of soil P availability cannot necessarily be translated to ecosystem processes elsewhere (6). I argue that the rise and decline of forest biomass is peculiar only to the biomes where regional tree-species diversity is impoverished. Unlike the Hawaiian model and the forest ecosystems studied in (1), conspicuous forest decline does not occur in the mainland tropics.

A meta-analysis of rain forests in the mainland tropics demonstrates that fresh-litter N:P ratios can vary widely, but that aboveground biomass does not drastically decline (Fig. 1). All sites far exceed the litter N:P ratio of 16 (the Redfield ratio), beyond which Wardle *et al.* argue that P is limiting biological processes relative to N. Aboveground biomass of mainland tropical rain forests increases with increasing leaf-litter N:P ratios for lowland sites or is more or less constant for montane sites; either case is inconsistent with the six chronosequences presented in (1). Moreover, Wardle *et al.* used basal area as an index of biomass, but this cannot be justified because tree height and not diameter is often a more decisive indicator of forest biomass.

In the case of Borneo, three tropical rain forests located approximately 1800 m above sea level (asl) and close to each other, but with contrasting soil P availability, do not drastically differ in forest biomass (7–9). Unlike a chronosequence, soil P availability in these forests differs as a result of geology. Nevertheless, the three sites form a gradient of soil P availability, which is in effect comparable to a chronosequence. In spite of six-fold differences in the pool of labile inorganic P and/or total P, the magnitude of the decline of forest biomass is small because of the displacements of P use by efficient species. In this soil P gradient, the fresh-litter N:P ratio more drastically increases with decreasing soil P pool than in the six chronosequences in (1).

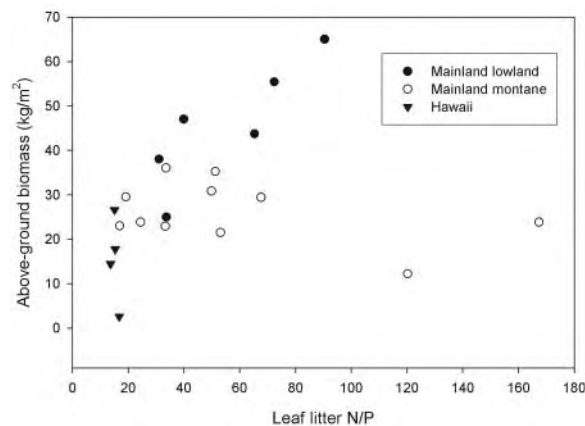


Fig. 1. N:P ratios of fresh leaf litter and aboveground biomass of rain forests in the mainland tropics (8, 10, 12–15) and in Hawaii (5). Tropical rain forests of the mainland are divided into lowland sites (below 1000 m asl) and montane sites (1000 to 2700 m asl) because of the possible interaction of air temperature and nutrients (P). For instance, plants need different amounts of P for a unit of photosynthesis under different air temperatures. The sites where both aboveground biomass and litter N:P are measured are included.

Why forest biomass does not drastically decline in the mainland tropics is an intriguing question. I argue that regional (not plot basis) tree-species diversity is two to three orders of magnitude greater in the mainland tropics than in the six chronosequences and that proportionately more tree species of high P-use efficiencies occur in the mainland tropics. Beta diversity of tree species across all six chronosequences is extremely low, which suggests that either the same taxon or relatively few taxa occur throughout each of the six chronosequences despite drastically different soil P pools [see species compositions in table S1 in (1)]. An extreme case is the Hawaiian chronosequence, where a single species, *Metrosideros polymorpha*, dominates the entire chronosequence. The ability of a single tree species to adapt to a wide range of phosphorus levels is limited relative to the adaptability of a collection of different species. The decline of forest biomass in response to reduced soil P availability is thus more dramatic in a monodominant system. In Southeast Asia, tropical rain forests can maintain an extremely large biomass (65 kg/m²) on infertile soils where litter N:P ratios exceed 90, which suggests a limitation of P relative to N (10). This clearly suggests that P-use-efficient species (11) maintain the biomass. Such plants probably use the labile P that is replenished from organic P on deeply weathered tropical soils. (6). The Wardle *et al.* study ignores the important role of biodiversity in terms of its community structure and function.

Kanehiro Kitayama

Center for Ecological Research

Kyoto University

509-3 Hirano 2-Chome, Otsu

Shiga 520-2113, Japan

E-mail: kitayama@ecology.kyoto-u.ac.jp

References

1. D. A. Wardle, L. R. Walker, R. D. Bardgett, *Science* **305**, 509 (2004).
2. T. W. Walker, J. K. Syers, *Geoderma* **15**, 1 (1976).
3. T. E. Crews *et al.*, *Ecology* **76**, 1407 (1995).
4. P. M. Vitousek, H. Farrington, *Biogeochemistry* **37**, 63 (1997).
5. D. A. Herbert, J. H. Fownes, *Ecosystems* **2**, 242 (1999).
6. A. H. Johnson, J. Frizano, D. R. Vann, *Oecologia* **135**, 487 (2003).
7. K. Kitayama, N. Majalap-Lee, S. Aiba, *Oecologia* **123**, 342 (2000).
8. K. Kitayama, S.-I. Aiba, *J. Ecol.* **90**, 37 (2002).
9. K. Kitayama, S.-I. Aiba, M. Takyu, N. Majalap, R. Wagai, *Ecosystems* **7**, 259 (2004).
10. J. Proctor, J. M. Anderson, P. Chai, H. W. Vallack, *J. Ecol.* **71**, 237 (1983).
11. P. M. Vitousek, *Ecology* **65**, 285 (1984).
12. P. J. Edwards, P. J. Grubb, *J. Ecol.* **65**, 943 (1977).
13. E. V. J. Tanner, *J. Ecol.* **65**, 883 (1977).
14. E. V. J. Tanner, *J. Ecol.* **68**, 573 (1980).
15. P. J. Edwards, *J. Ecol.* **70**, 807 (1982).

7 January 2005; accepted 21 March 2005
10.1126/science.1109537

Response to Comment on “Ecosystem Properties and Forest Decline in Contrasting Long-Term Chronosequences”

Kitayama (1) correctly recognizes that the six forested chronosequences we studied (2) do not include any of the hyperdiverse forests commonly found in the tropics. We welcome others to test whether the patterns we found for our six sites also occur in other systems and also recognize that there are plausible reasons as to why these patterns may or may not occur. However, we believe that the way Kitayama has attempted to test the validity of our ideas with regard to hyperdiverse tropical rainforests has substantial limitations.

First, Kitayama pools data from six studies spanning the continental tropics and suggests that nitrogen to phosphorus (N:P) ratios and aboveground biomass are not negatively related. However, this analysis involves individual sites that span very large spatial scales, across which a range of other driving factors (notably macroclimate, geology, and disturbance regime) would vary considerably and could well override effects of soil fertility. It is well recognized that macroclimate is the key driver of biogeochemical processes and ecosystem functioning at large spatial scales, whereas variables related to resource quality are more important at local spatial scales, where variations in climate are more likely to be minor (3, 4). Kitayama's demonstration that N:P ratios are unimportant as drivers of tree biomass at large spatial scales therefore does not preclude these ratios from being important drivers at local spatial scales such as at the within-chronosequence scale that we studied (2). His analysis is therefore not relevant to the question that we have addressed. Kitayama's own work in Borneo (5) is more relevant, but is based on only three sites that have formed on different geological substrates rather than by pedogenesis along a single substrate, as in our work. Whether his findings are characteristic of hyperdiverse tropical forests at large remains to be

tested, and we maintain that this question could best be tested by using long-term chronosequence data.

Second, we have particular concerns about the mechanistic basis that Kitayama proposes to explain why hyperdiverse tropical rainforests apparently do not decline drastically with increasing P limitation, i.e., that the additional diversity that these forests have over less diverse ones confers benefits for forest stand nutrient efficiency. We note that our chronosequences are not “extremely low” and “impoverished” in diversity, as claimed by Kitayama; the Franz Josef and Cooloola sequences in particular have more than 30 tree species that occur with some abundance. More important, Kitayama's proposed diversity mechanism requires that increasing tree diversity from tens of species (our study sites) to hundreds of species (hyperdiverse tropical rainforests) has beneficial effects on ecosystem functions (in this case, the ability of forest stands to resist decline by P limitation). This mechanism assumes that ecosystem functioning (i.e., the rate of ecosystem-level processes) increases monotonically with increasing diversity at levels of plant species richness well beyond 100 species. However, several studies have found little effect of plant diversity on ecosystem functioning except at very low levels of diversity (6, 7). Even those studies that have been the most generous in ascribing positive roles of biodiversity to ecosystem functioning (8, 9) have suggested an asymptotic relation between diversity and function, which effectively saturates at a diversity of around 10 species. This relation is also supported by theoretical predictions (10). The mechanistic basis that Kitayama proposes therefore has no support from literature on the diversity-function issue, whether theoretical, experimental, or empirical.

Finally, Kitayama claims that our use of basal area to show biomass decline during

retrogression is not justified. However, this basal-area decline is matched by published measured declines of tree height and/or tree biomass for most of these chronosequences (11–14). It is indisputable that tree biomass drops sharply during retrogression for each of our chronosequences, and this is apparent from visual inspection of the stands that we considered, including photographs of them [see figure S1 in (2)].

D. A. Wardle

*Department of Forest Vegetation Ecology
Swedish University of Agricultural Sciences
SE901-83 Umeå, Sweden*

*E-mail: david.wardle@svek.slu.se
and Landcare Research,
Post Office Box 69,
Lincoln, New Zealand*

L. R. Walker

*Department of Biological Sciences
Box 454004*

*University of Nevada, Las Vegas
4505 Maryland Parkway
Las Vegas, NV 89154-4004, USA*

R. D. Bardgett

*Institute of Environmental and
Natural Sciences
University of Lancaster
Lancaster LA1 4YQ, UK*

References and Notes

1. K. Kitayama, *Science* **308**, 633 (2005); www.sciencemag.org/cgi/content/full/308/5722/633b.
2. D. A. Wardle, L. R. Walker, R. D. Bardgett, *Science* **305**, 509 (2004).
3. M. J. Swift, O. W. Heal, J. M. Anderson, *Decomposition and Terrestrial Ecosystems* (Blackwell, Oxford, 1979).
4. M. M. Couteaux, P. Bottner, B. Berg, *Trends Ecol. Evol.* **10**, 63 (1995).
5. K. Kitayama, S.-I. Aiba, M. Takyu, N. Majalap, R. Wagai, *Ecosystems* **7**, 259 (2004).
6. D. U. Hooper, P. M. Vitousek, *Ecol. Monogr.* **68**, 121 (1998).
7. D. A. Wardle, O. Zackrisson, G. Hörnberg, C. Gallet, *Science* **277**, 1296 (1997).
8. D. Tilman, D. Wedin, J. Knops, *Nature* **379**, 718 (1996).
9. A. Hector *et al.*, *Science* **286**, 1123 (1999).
10. M. Schwartz *et al.*, *Oecologia* **122**, 297 (2000).
11. T. E. Crews *et al.*, *Ecology* **76**, 1407 (1995).
12. S. Richardson, D. A. Peltzer, R. B. Allen, M. S. McGlone, R. L. Parfitt, *Oecologia* **139**, 267 (2004).
13. D. A. Wardle, G. Hörnberg, O. Zackrisson, M. Kalela-Brundin, D. Coomes, *Science* **300**, 972 (2003).
14. A. F. Mark, C. Grealish, C. M. Ward, J. B. Wilson, *J. R. Soc. N.Z.* **18**, 29 (1983).
15. We thank R. Allen, P. Bellingham, and D. Peltzer for helpful discussions during the preparation of this response.

21 January 2005; accepted 25 March 2005
10.1126/science/1109723

Ideals Drowned in the Marketplace

Harold T. Shapiro

Jennifer Washburn has written a very engaging and useful book. Moreover, *University, Inc.* is a book with attitude.

The author focuses her attention and dismay on the relentless growth in the commercialization of education and scholarship within universities, the resultant diminishment of the intellectual commons, and the

elimination or blurring of the boundaries that separate the distinct values and virtues of academic life from those of the commercial sector. Washburn's quite legitimate concern, which I share, is that the university's roles as a disinterested arbiter of knowledge and society's thoughtful

critic could be lost in an environment that is increasingly dominated by the university's unrestrained search for increased resources.

The book has both pluses and minuses, but I am glad it has been written. In addition, I would encourage all those concerned with the future of American higher education to read it. Although the style and approach will please some and alienate others, Washburn thoughtfully addresses a set of important issues that are often absent from our national discourse on higher education.

The book's most important asset is the author's genuine concern with the values that ought to help structure the role and the function of universities in a society where the influence of private markets and their associated materialistic values is ever more pervasive. Washburn offers a series of well-researched cases that, at the very least, are sobering reminders that things can and will go wrong unless universities pay more than lip service to their values and their distinct role in society. Thus, despite the author's penchant for overstatement and for the substitution of anecdotes for data and colorful phrases for thoughtful assessments, she does lay bare actual cases where highly respected universities and their faculties have behaved hypocritically and abandoned important academic norms in the service of accumulating more

resources. As Derek Bok (former president of Harvard University) and others have pointed out, if universities cannot sustain a set of values that define what they will not do even for money, there is little chance that they will fulfill their social responsibilities.

The author, a journalist and fellow at the New America Foundation, also provides a useful historical sketch of how all this has come about. Her account ranges from Thomas Jefferson's view of the "utilitarian university," through the 19th-century controversy over whether higher education should stress utilitarian goals or those of liberal learning, to the impact of the Bayh-Dole Act of 1980. Nevertheless, I find this aspect of the book long on nostalgia for the good old days. Washburn seems not to recognize that in those days American colleges and universities fulfilled neither their scholarly nor educational functions and were largely irrelevant to the nation's affairs, public or private.

Even more perplexing is that Washburn gives little thoughtful attention to certain critically important contemporary circumstances that directly impact and constrain the evolving role of academic institutions in the United States. For example, developments both on scientific frontiers and in financial markets are changing the balance of forces between for-profit and not-for-profit institutions within the scientific endeavor. As a result, not only has industry's share of the research enterprise grown much more quickly than the corresponding academic-based share, but the nature of certain academic-based research efforts has changed. Besides, it is not sufficient to simply take passing notice of the steadily declining state support for U.S. public universities; one must consider just how these vital universities are supposed to sustain their quality in such an environment.

Nor does Washburn give serious consideration to the fact that the evolving tapestry of American higher education involves many quite distinct narratives. Indeed, given the het-

erogeneity of American higher education, its civic purposes must be understood as requiring different responses from differently situated institutions. Whereas the author seems uncomfortable with the decentralization of American higher education, I think that feature is one of its glories. Yes, the freedoms of decentralization produce excesses, just as personal freedoms do. (In particular, it may be harder for academic institutions acting alone to resist the corrosive aspects of the increasing commercialization of their activities.) Nonetheless, the benefits offered by such freedoms, for both persons and institutions, more than compensate for the costs that are incurred by the excesses that do occur. Lastly, given that the social responsibilities of American universities cover such a wide spectrum of activities, the book reflects little understanding of exactly how variously situated universities can locate a position of dynamic equipoise among the many different legitimate responsibilities they have.

The book concludes with suggestions for a path forward. Washburn calls for universities to make a renewed commitment to academic values—a call I heartily second. She also offers a set of recommendations that generally increase the authority of the federal government or other third parties. With the exception of a proposal aimed at strengthening conflict-of-interest regulations, I find these unpersuasive and, in many cases, a little naïve.

Despite these shortcomings, *University, Inc.* is well worth reading. Washburn's mostly fair-minded and engaging presentation of various examples (principally from biomedicine) reminds us that in a rapidly changing world we need to be constantly and thoughtfully re-viewing whether the nature and content of higher education's portfolio of activities and policies need to be rebalanced or redirected to help universities fulfill their highest social role. It is not that academic values need to remain fixed; indeed, universities must both create and adapt to change. But to do so in ways that best fulfill their social responsibilities, they need to know which kinds of activities they must do, which they may do, and which they will not do if they are to avoid undermining their most important values.

Furthermore, Washburn is correct on a crucially important point: the values of the market can and will overtake those of aca-

**University, Inc.
The Corporate
Corruption of
American Higher
Education**

by Jennifer Washburn
Basic Books, New York,
2005. 346 pp. \$26,
C\$36.95, £19.99. ISBN
0-465-09051-6.



The reviewer is the emeritus president of Princeton University, 355 Wallace Hall, Princeton, NJ 08544, USA. E-mail: hts@princeton.edu

demia unless universities guard the latter carefully. We must hope that universities will do so. Society developed a separate set of institutional frameworks for not-for-profit organizations because many important social objectives cannot be realized through the operation of private markets. Despite the great strengths of market mechanisms, they cannot provide all of the goods and services society needs.

10.1126/science.1110892

MICROBIOLOGY

Appreciating Our Usual Guests

Elaine Tuomanen

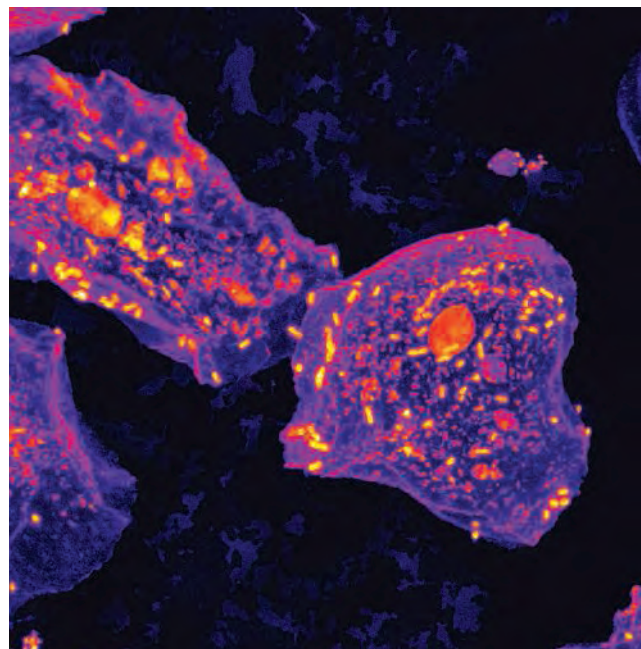
Mention bacteria mingling with humans and the first response is alarm. In fact, the vast majority of attention paid to the bacterial-human interaction focuses on pathogenesis. In *Microbial Inhabitants of Humans*, Michael Wilson cogently reminds us that by far the more representative interactions are not pathological but symbiotic. This authoritative book fills a gaping hole in the literature by describing how bacteria interrelate with us. Wilson's basic premise is that both humans and bacteria benefit from peaceful coexistence and that microbial colonization plays an obligatory role in human health (as much as invasion plays in disease). The book's strength lies in the exhaustive, clearly organized evidence that the author presents to overwhelmingly document his hypothesis. Wilson (a microbiologist at the Eastman Dental Institute, University College London) even ventures into the area of manipulation of colonization by probiotics, prebiotics, and replacement therapies. His hard-nosed look at the evidence in the area of holistic medicine will be much appreciated by the skeptic and the enthusiast alike.

Some astounding numbers justify attention to our bacterial symbionts. There are ten times more bacteria colonizing a human than the number of human cells in the body (10^{14} versus 10^{13} , respectively). Over 700 taxa can be found at a single site. The structures of communities vary tremendously. The gut might be considered New York

City, whereas the skin is perhaps more like Memphis. Over 30 years ago, dentists brought attention to the concept of polymicrobial communities in the formation of dental caries. Seeking interventions, they began serious work on the molecular mechanisms of bacterial attachment to host cells, tooth components, and even other bacteria. The field then grew tremendously as its implications for mechanisms of pathogenesis in a broad range of infections became clear. Now we are familiar with microcolonies that become biofilms, the complex, matrix-enclosed ecosystems described structurally by fractals. Even more recently, new podlike structures built by host-bacterial interactions have been discovered in the urinary tract. The communication between colonizing microbes and host innate defense that drives the evolution of these fascinating niches relies heavily on quorum sensing. Bacteria regulate their community activities using secreted peptides or small molecules, and hosts detect or interfere with this cross-talk to either cooperate with or kill the newcomers. That this process is often peaceful and mutually beneficial is a central message of the book.

The chapters cover the detailed relevant anatomy and local defenses for each colonized site: skin, eye, respiratory tract, genitourinary tract, gastrointestinal tract, and mouth. Nonmedical readers will find this information particularly valuable for setting the architecture that underlies the symbiotic interactions. For each anatomical site, Wilson provides a list of which bacteria are found where along with brief descriptions of how the community is sustained in healthy humans and how it is disrupted during disease. The author also points out changes in the biology of the community as a function of age of the host, production of hormones, etc. He has made a valuable contribution by assembling in one place this comprehensive body of information.

Wilson interprets the patterns he documents as evidence that indigenous microbes help maintain human health. For instance, resident bacteria will kill incoming bacteria and thereby protect the host. The author presents a fascinating table that compares features of germ-free and normal animals. Among the differences: the presence of res-



Hard to hold on. Because mechanical and hydrodynamic forces tend to dislodge microbial aggregates, many epithelial cells—such as these from the cheek mucosa viewed in confocal laser scanning microscopy—have only small numbers of individual bacteria attached to their surfaces.

ident microbiota enhances organ size, nutrition (e.g., provides vitamins), detoxification of ingested toxins, and chemotaxis of macrophages and lymphocytes.

The author explores the therapeutic value of manipulating the indigenous microbiota in a well-balanced discussion of probiotics that considers the use of nonpermanent colonizers for treatment of diarrhea, vaginosis, and dental cavities. He briefly discusses prebiotics, nondigestible foods that benefit the host by promoting one or a few bacteria species in the colon. And he presents the inhibition of microbial adhesion as a possible mode of therapy for tooth decay, urinary tract infections (e.g., cranberry juice contains an effective anti-adhesive for *Escherichia coli*), and infections associated with the use of medical devices. As in earlier chapters, he provides tables that summarize scientifically valid studies for each area.

Despite the many benefits of indigenous microbiota, disease can clearly arise from the “friendlies” in the context of abnormalities in the host response (such as immunocompromise) or introduction of agents that break the peace (for instance, catheters). Nonetheless, *Microbial Inhabitants of Humans* is striking for its balanced view of the data. Rather than the usual litany of threats, Wilson offers a more appropriate acknowledgment of the benefits attending coexistence with our indigenous microbes.

**Microbial Inhabitants of Humans
Their Ecology and Role in Health and Disease
by Michael Wilson**

Cambridge University Press,
New York, 2005. 475 pp. \$65,
£35. ISBN 0-521-84158-5.

The reviewer is at the Department of Infectious Diseases, St. Jude Children's Research Hospital, Mailstop 320, 332 North Lauderdale Road, Memphis, TN 38105, USA. E-mail: elaine.tuomanen@stjude.org

10.1126/science.1111460

Synthesizing U.S. River Restoration Efforts

E. S. Bernhardt,^{1*} M. A. Palmer,¹ J. D. Allan,² G. Alexander,² K. Barnas,³ S. Brooks,⁴ J. Carr,⁵ S. Clayton,⁶ C. Dahm,⁷ J. Follstad-Shah,⁷ D. Galat,^{8,9} S. Gloss,¹⁰ P. Goodwin,⁶ D. Hart,⁵ B. Hassett,¹ R. Jenkinson,¹¹ S. Katz,³ G. M. Kondolf,¹² P. S. Lake,⁴ R. Lave,¹² J. L. Meyer,¹³ T. K. O'Donnell,⁹ L. Pagano,¹² B. Powell,¹⁴ E. Sudduth¹³

The importance of rivers and streams for fresh water, food, and recreation is well known, yet there is increasing evidence that degradation of running waters is at an all-time high (1). More than one-third of the rivers in the United States are listed as impaired or polluted (2), and freshwater withdrawals in some regions are so extreme that some major rivers no longer flow to the sea year round (3). Extinction rates of freshwater fauna are five times that for terrestrial biota (4, 5). Fortunately, stream and river restoration can lead to species recovery, improved inland and coastal water quality, and new areas for wildlife habitat and recreational activities (6–11).

River restoration has become a highly profitable business (12, 13) and will play an increasing role in environmental management and policy decisions (7). A few high-profile and large restoration projects such as those on the Kissimmee River (11, 14) and the Grand Canyon (15, 16) are well documented. However, most restoration projects are small scale (implemented on less than 1 km of stream length), and information on their implementation and outcome is not readily accessible. This prompted us to build a database of river restoration across the United States with the goal of determining the common elements of successful projects.

We found that existing restoration databases are highly fragmented and often rely on ad hoc or volunteer data entry. Thus, we developed methods for the unbiased collection and cataloging of river and stream restoration projects. Here, we report a synthesis of information on 37,099 projects in the National River Restoration Science Synthesis (NRRSS) database.

The NRRSS database includes all stream and river restoration projects present in national databases as of July 2004, as well as a large sample of river and stream restoration projects from seven geographic regions (see figure, below) [(17) part a]. Because we wanted to document how restoration dollars and efforts were allocated, we did not limit data collection to projects that fit our definition of restoration. No judgments were made of the validity of the terms “stream restoration” or “project.” Use of national coverage data sources [(17) part b] ensured inclusion of projects from all 50 states. For the seven specific regions, we also collected information on all restoration projects for which we could obtain data, regardless of project size, restoration method, implementer, or perceived suc-

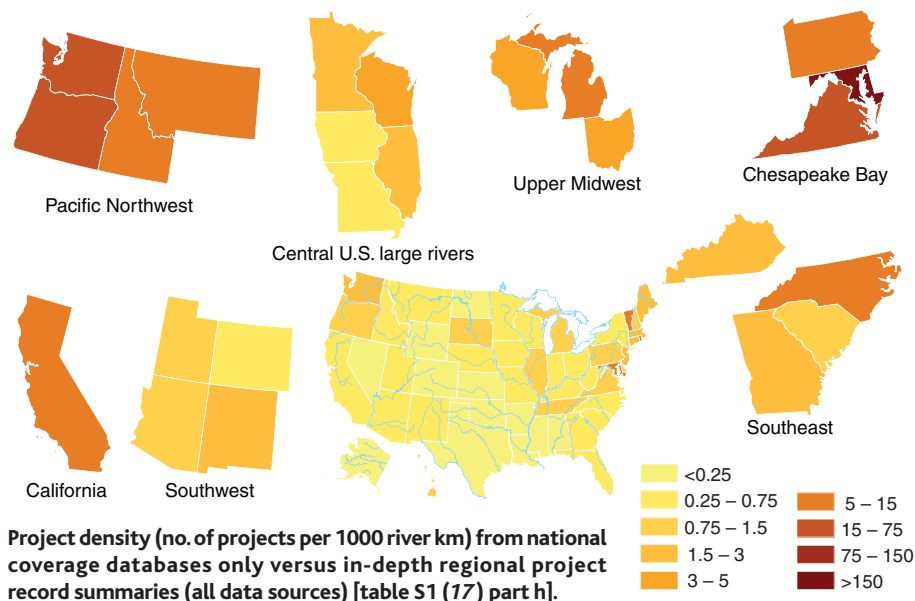
cess or failure of the project. We identified a priori 13 categories of restoration and classified each project according to its stated goal [see table, page 637 and (17) part c].

The number of river restoration projects increased exponentially during the last decade, paralleling the increase in news media and scientific reports [fig. S1 (17) part d]. However, restoration efforts varied across geographic regions. Most projects (88%) are from the Pacific Northwest, the Chesapeake Bay watershed, or California (see figure, below). Data from national coverage sources [(17) part b] made up <8% of projects in the NRRSS database. Thus, while federal funding supports some tracking efforts, national restoration databases are not tracking the majority of projects and lack information on the regional differences in expenditures and effort found with our approach.

The most commonly stated goals for river restoration in the United States are (i) to enhance water quality, (ii) to manage riparian zones, (iii) to improve in-stream habitat, (iv) for fish passage, and (v) for bank stabilization (see figure, page 637). Projects with these goals are typically small in scale with median costs of <\$45K (see table, page 637). A large proportion of restoration dollars are spent on fewer, more expensive projects aimed at reconnecting floodplains, modifying flows, improving aesthetics or recreation, and reconfiguring river and stream channels (see figure, page 637). Of the projects in our database, 20% had no listed goals; in many cases, descriptions were too limited to determine whether projects were undertaken to restore stream ecosystems or were merely river manipulation projects (e.g., bank stabilization) (18).

¹University of Maryland, College Park, MD; ²School of Natural Resources and Environment, University of Michigan, Ann Arbor, MI; ³Northwest Fisheries Science Center (NWFS), National Oceanographic and Space Administration (NOAA), Seattle, WA, USA. ⁴Monash University, Clayton, VIC, Australia. ⁵Academy of Natural Sciences, Philadelphia, PA; ⁶University of Idaho, Boise, ID; ⁷University of New Mexico, Albuquerque, NM; ⁸U.S. Geological Survey (USGS), Cooperative Research Units, Columbia, MO; ⁹University of Missouri, Columbia, MO; ¹⁰Southwest Biological Science Center, USGS, Tucson, AZ; ¹¹University of Idaho, Moscow, ID; ¹²University of California, Berkeley, CA; ¹³University of Georgia, Athens, GA; ¹⁴Center for Biological Informatics, USGS, Denver, CO, USA. Complete addresses are available online.

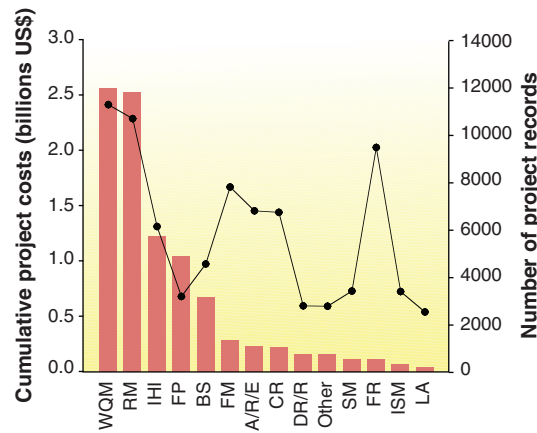
*Present address: Duke University, Durham, NC, USA.
[†]Author for correspondence. E-mail: emily.bernhardt@duke.edu



Only 58% of the project records used to populate our database had information on project costs. For this subset, total costs came to \$9.1 billion. Most of this was spent after 1990, with \$7.5 billion in recorded costs from 1990 to 2003 (from the 58% reporting costs). Applying this cost estimate to the remaining ~40% of projects [(17) part e], and taking into account that we captured ~27% of all stream and river restoration projects in the 27 states not within one of our regional nodes [(17) part e], at least \$14 to \$15 billion has been spent on restoration of streams and rivers within the continental United States since 1990, an average of >\$1 billion a year. This is probably an underestimate, because data providers reported that the costs listed in project records typically do not include matching or in-kind contributions such as agency labor. In addition, the data sources we accessed did not capture costs for the restoration of the Kissimmee River or the full costs of Glen Canyon, San Francisco Bay, Columbia, and Missouri river restoration efforts, which would add hundreds of millions to billions of dollars (17).

Our analysis confirms what the General Accounting Office (GAO) has suggested in recent reports to the U.S. Congress (19, 20): a comprehensive assessment of restoration progress for the United States, or even for individual regions, is not possible with the “piecemeal” information currently available. We found that only 10% of project records indicated that any form of assessment or monitoring occurred. Most of these ~3700 projects were not designed to evaluate consequences of restoration activities or to disseminate monitoring results.

Monitoring and assessment varied by



Distribution of projects within each restoration goal category. Abbreviations of categories are in table below.

region: >20% of projects in the Southwest, Southeast, and Central United States had some form of monitoring, whereas only 6% of project records in the Chesapeake Bay watershed indicated that monitoring occurred (see figure, page 636). Projects with higher costs were more likely to be monitored [average costs were $\$1.5 \pm \0.7 million (95% CI), whereas unmonitored project costs were $\$0.4 \pm \0.08 million]. Regions with greater project density tended to have lower average project costs and reported a lower rate of monitoring. Further, differences in regional regulations are likely.

Because most project records were inadequate to extract even the most rudimentary information on project actions and outcomes, it is apparent that many opportunities to learn from successes and failures, and thus to improve future practice, are being lost. The largest and most costly programs have recognized this problem and have enacted solutions (16, 19). Unfortunately, the outcomes of most of the tens of thousands of projects of small-to-modest size are currently not adequately

tracked, yet cumulatively, their costs are greater, and their reach is far broader. Much greater effort is needed to gather and disseminate data on restoration methods and outcomes, particularly given the magnitude of costs. It is unrealistic to expect that every restoration project will have extensive monitoring activities, but strategic pre- and postassessments with standardized methods could enable restoration practitioners and managers to understand what types of activity are accomplishing their goals (21). Ensuring data compatibility in the tracking of restoration projects and the documentation of results from project evaluations are equally important. To facilitate this effort, the NRRSS database structure and schema are freely available (22).

References and Notes

1. P. Gleick, *Science* **302**, 1524 (2003).
2. U.S. Environmental Protection Agency (EPA), “National Water Quality Inventory” (EPA Publ. 841-R-02-001, Washington, DC, 2000).
3. National Research Council (NRC), *Our Common Journey: A Transition Toward Sustainability* (National Academy Press, Washington, DC, 1999).
4. K. Sand-Jensen, in *Encyclopedia of Biodiversity* (Academic Press, San Diego, 2001), pp. 89–108.
5. A. Ricciardi, J. B. Rasmussen, *Conserv. Biol.* **13**, 1220 (1999).
6. J. S. Baron et al., *Ecol. Appl.* **12**, 1247 (2002).
7. M. Palmer et al., *Science* **304**, 1251 (2004).
8. S. Postel, B. Richter, *Rivers for Life: Managing Waters for People and Nature* (Island Press, Washington, DC, 2003).
9. NRC, *Upstream: Salmon and Society in the Pacific Northwest* (National Academy Press, Washington, DC, 1996).
10. A. D. Buijse et al., *Freshwater Biol.* **47**, 889 (2002).
11. P. J. Whalen et al., *Water Sci. Technol.* **45**, 55 (2002).
12. B. Lavendel, *Ecol. Restor.* **20**, 173 (2002).
13. D. Malakoff, *Science* **305**, 937 (2004).
14. J. W. Koebel, *Restor. Ecol.* **3**, 149 (1995).
15. J. P. Cohn, *Bioscience* **51**, 998 (2001).
16. NRC, *Downstream: Adaptive Management of Glen Canyon Dam and the Colorado River Ecosystem* (National Academy Press, Washington, DC, 1999).
17. See Supporting Online Material.
18. S. Gillilan et al., *J. Appl. Ecol.*, in press.
19. “Columbia River Basin salmon and steelhead: Federal agencies’ recovery responsibilities, expenditures, and actions,” *GAO Tech. Rep. GAO-02-612* (2002).
20. “Great Lakes: An overall strategy and indicators for measuring progress are needed to better achieve restoration goals,” *GAO Tech. Rep. GAO-03-515* (2003).
21. M. A. Palmer et al., *J. Appl. Ecol.*, in press.
22. <http://nrrss.nbio.gov>
23. The NRRSS is supported by the National Center for Ecological Analysis and Synthesis (NCEAS). The national effort received support from NSF, USGS, the David and Lucille Packard Foundation, Altria, and the U.S. EPA. Individual regional teams received support from the C. S. Mott Foundation, the CRC for Freshwater Ecology, the McKnight Foundation, CalFed, the U.S. Bureau of Reclamation, University of Maryland, and the Maryland DNR. USGS National Biological Information Infrastructure (NBII) provided personnel support; we particularly thank G. Cotter and M. Frame. We thank all data providers (17), particularly NWFSC NOAA, M. Ehrhart, S. D. Kunkoski, M. Wiley, and P. Steen; also R. Carlson and K. Ward who provided us with previously synthesized regional databases. Views expressed here do not represent the views of any supporting organization or data provider.

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5722/636/DC1

10.1126/science.1109769

MEDIAN COSTS FOR GOAL CATEGORIES

NRRSS goal category	Median cost	Examples of common restoration activities
Aesthetics/recreation/education (A/R/E)	\$63,000	Cleaning (e.g., trash removal)
Bank stabilization (BS)	\$42,000	Revegetation, bank grading
Channel reconfiguration (CR)	\$120,000	Bank or channel reshaping
Dam removal/retrofit (DR/R)	\$98,000	Revegetation
Fish passage (FP)	\$30,000	Fish ladders installed
Floodplain reconnection (FR)	\$207,000	Bank or channel reshaping
Flow modification (FM)	\$198,000	Flow regime enhancement
Instream habitat improvement (IH/I)	\$20,000	Boulders/woody debris added
Instream species management (ISM)	\$77,000	Native species reintroduction
Land acquisition (LA)	\$812,000	
Riparian management (RM)	\$15,000	Livestock exclusion
Stormwater management (SM)	\$180,000	Wetland construction
Water quality management (WQM)	\$19,000	Riparian buffer creation/maintenance

Median costs for goal categories.

Toward Bridging the Terahertz Gap with Silicon-Based Lasers

Alexander Borak

Many technologies use the optical and electronic regions of the electromagnetic spectrum. Wavelengths $<30\ \mu\text{m}$ are used in optical data storage, fiber-optic communications, and spectroscopy, whereas wavelengths $>300\ \mu\text{m}$ are the domain of electronics, radio communications, and radar. In contrast, the “terahertz gap” from 30 to $300\ \mu\text{m}$ (1 to 10 THz) has barely been exploited because no cheap and practical sources of terahertz radiation exist. In recent years, substantial progress has been made in developing such devices. Terahertz lasers based on silicon would be particularly desirable because of their compatibility with silicon technology.

The development of a silicon-based terahertz laser is part of the larger effort to build the first silicon-based laser. The latter effort recently received a boost with the announcement of an optically pumped silicon Raman laser operating at $\sim 1.7\ \mu\text{m}$ (1). Such near-infrared wavelengths are important for telecommunications.

A cheap and compact source of terahertz radiation would open up many exciting applications (2, 3). Many complex molecules have rotational and vibrational modes in this region, and many materials such as plastics, clothing, and semiconductors are transparent to terahertz radiation. Terahertz spectroscopy could therefore be used to detect and identify explosives, bioweapons, and narcotics, as well as in cancer screening and proteomics (the study of proteins within the body). Terahertz radiation can also be used for structural imaging, much like x-rays but with greater capabilities. For example, because terahertz radiation is readily absorbed by water, terahertz imaging could be used in dental or skin cancer imaging to differentiate between different tissue types. Security screening with terahertz imaging could detect ceramics as well as metals. Terahertz imaging would also allow nondestructive testing of a wide range of products in production monitoring.

lasing at similar wavelengths from silicon monocystals doped with arsenic (8).

These silicon-based terahertz lasers are being developed for applications in astronomy and atmospheric spectroscopy. However, they require a separate optical pump laser and operate only at low temperatures. An ideal laser would be electrically pumped, directly emit terahertz radiation, and operate at room temperature.

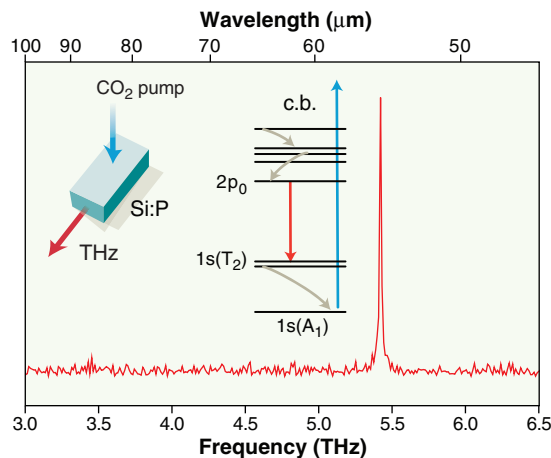
Impurity emission from doped silicon can also be initiated by electrical pumping. Lv *et al.* recently demonstrated impurity-related electroluminescence from 20 to $50\ \mu\text{m}$ (6 to 14 THz) from doped silicon under pulsed currents (9). The results are encouraging and the devices easy to fabricate, but the emission degrades above 20 K. Furthermore, if the devices are to lase, population inversion is required, and no evidence of this was seen for these structures.

In 2002, Köhler *et al.* reported the development of a terahertz quantum cascade laser based on layered III-V semiconductor heterostructures (10). The work was a major step toward a compact and practical electrically pumped terahertz emitter. In a quantum cascade laser, the charge carriers cascade through a series of layers (“quantum wells”) while emitting multiple photons (see the second figure). Population inversion is achieved through careful control of the lifetimes of the upper and lower states. Quantum cascade lasers offer high

output powers and control of layer thickness, allowing the emission wavelength to be designed in. The latest III-V terahertz quantum cascade lasers produce laser radiation at 2.9 THz with a power of 15 mW at 10 K (11).

However, III-V semiconductors are polar materials, and photons with energies below the optical phonon energy ($\sim 36\ \text{meV}$ for GaAs) therefore experience strong polar optical phonon scattering. This effect reduces the lifetime of the upper state as the temperature increases and is expected to severely limit the high-temperature operation of terahertz quantum cascade lasers. Indeed, the highest operating temperature achieved to date with a III-V terahertz quantum cascade laser in pulsed operation is only $\sim 150\ \text{K}$ (12).

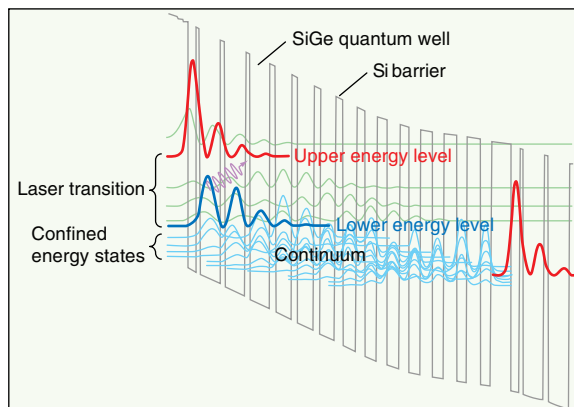
Quantum cascade lasers based on silicon/germanium heterostructures are not expected to suffer from this limitation, because silicon is nonpolar and the silicon-germanium bond results in negligible polar



A terahertz phosphorus-doped silicon (Si:P) laser. A sharp emission is seen at $\sim 5.5\ \text{THz}$. Such terahertz silicon lasers may eventually be used on telescopes such as the Stratospheric Observatory for Infrared Astronomy. **Left inset:** Optical pumping of Si:P (blue arrow) leads to terahertz laser emission (red arrow). **Middle inset:** In Si:P, pump photons excite carriers from the $1s$ ground state to the conduction band (c.b.) (blue arrow). The carriers fall to the $2p_0$ state by nonoptical transitions (gray arrows). During the optical transition between the $2p_0$ and $1s$ states (red arrow), a photon is emitted. Carriers then return to the $1s$ state. $1s(A_1)$, $1s(T_2)$, and $2p_0$ are the atomic-like states of the phosphorus dopant atoms.

Current terahertz lasers typically use a mid-infrared CO_2 laser to optically excite the molecules in a gas cell (4). However, such systems are complex and expensive. A more recent development relies on transitions between impurity states in doped silicon, thus replacing the gas cell with a more practical semiconductor crystal. Hübers and co-workers have demonstrated lasing at wavelengths of 50 to $60\ \mu\text{m}$ (5 to 6 THz) from bulk silicon doped with phosphorus, bismuth, and antimony at temperatures of up to 30 K (see the first figure) (5–7). In these systems, the emitted photon energy depends on the energy spacing of the impurity states and hence on the dopant species used. The operation of any laser depends on achieving population inversion, where there are more carriers in an upper energy level than in a lower one. In doped bulk silicon, this can occur as a result of the relatively long lifetime of the excited state. Recently, Hübers *et al.* have reported

The author is at ICI Strategic Technology Group, Wilton Centre, Wilton, Redcar TS10 4RF, UK. E-mail: alex_borak@ici.com



optical phonon scattering. In III-V terahertz quantum cascade lasers, the upper state lifetime is substantially reduced above 40 K, but in silicon/germanium structures, time-resolved experiments have shown constant lifetimes up to ~ 150 K (13). Silicon also has a higher thermal conductivity than III-V materials. A silicon-based quantum cascade laser therefore promises to be a good candidate for a room-temperature terahertz source.

Because of material considerations, all silicon/germanium quantum cascade structures investigated to date have been based on transitions in the valence band. Unfortunately, the valence band is made up of many interacting subbands, and the carriers are holes (as opposed to electrons in the

One period of a silicon/germanium quantum cascade laser. Traveling from left to right, the carriers enter the upper energy level, emit a photon upon falling to the lower level, and then move rapidly through the continuum to be reinjected into the upper level of the next period. A terahertz quantum cascade laser can have more than 100 such periods. Population inversion is achieved by designing the upper level to have a longer lifetime than the lower one, which is rapidly depopulated by the continuum.

conduction band) with a very high effective mass. These and other factors make the design of successful silicon/germanium quantum cascade structures more challenging than is the case for III-V materials.

Lynch *et al.* demonstrated electroluminescence at 2.9 THz from transitions between energy levels in the same well (14). Bates *et al.* obtained similar results at 1.2 THz from transitions between energy levels in neighboring wells; such interwell transitions promise an increased upper state lifetime (15). Recently, Paul *et al.* have grown a cascade structure with a buried tungsten silicide layer (16). Such silicides may provide the means to grow cladding layers with good electrical conductivity but low optical absorption, vital for successful laser operation.

Optically pumped silicon impurity lasers in the terahertz range have been around for some years (5–8), but a compact, electrically pumped terahertz laser operating at room temperature remains elusive.

The quantum cascade approach is arguably the most promising; here, silicon/germanium structures may offer key advantages over III-V materials for high-temperature operation. However, serious obstacles must be overcome before a working silicon quantum cascade laser can be produced.

References and Notes

1. H. Rong *et al.*, *Nature* **433**, 725 (2005).
2. D. D. Arnone *et al.*, *Phys. World* **13** (4), 35 (2000).
3. TeraView Ltd. (www.teraview.co.uk).
4. E. R. Mueller, *Industrial Physicist*, 27 (August/September 2003).
5. S. G. Pavlov *et al.*, *Phys. Rev. Lett.* **84**, 5220 (2000).
6. S. G. Pavlov *et al.*, *Appl. Phys. Lett.* **80**, 4717 (2002).
7. S. G. Pavlov *et al.*, *J. Appl. Phys.* **92**, 5632 (2002).
8. H.-W. Hübers *et al.*, *Appl. Phys. Lett.* **84**, 3600 (2004).
9. P. Lv *et al.*, *MRS Symp. Proc.* **832**, F4.3.1-9 (2005).
10. R. Köhler *et al.*, *Nature* **417**, 156 (2002).
11. S. Barbieri *et al.*, *Appl. Phys. Lett.* **85**, 1674 (2004).
12. B. S. Williams *et al.*, *Appl. Phys. Lett.* **83**, 5142 (2003).
13. P. Murzyn *et al.*, *Appl. Phys. Lett.* **80**, 1456 (2002).
14. S. A. Lynch *et al.*, *Appl. Phys. Lett.* **81**, 1543 (2002).
15. R. Bates *et al.*, *Appl. Phys. Lett.* **83**, 4092 (2003).
16. D. J. Paul *et al.*, *MRS Symp. Proc.* **832**, F4.1.1-9 (2005).
17. During preparation of this Perspective, the author was at the Laboratory for Micro and Nanotechnology, Paul Scherrer Institut, 5232 Villigen PSI, Switzerland.

10.1126/science.1109831

SOCIOLOGY

Network Theory—the Emergence of the Creative Enterprise

Albert-László Barabási

In the *Foundation Trilogy*, Isaac Asimov placed psychohistorian Harry Seldon so far into the future that Earth, the birthplace of the Galactic civilization, has been forgotten (1). Indeed, Star Trek's teleporting characters appear far more grounded in reality than Seldon's mathematical equations that accurately predict the multigalactic society's fate thousands of years into the future. Today, when reports about quantum teleportation fill the pages of the best physics journals, we wonder how long it will be until a real Harry Seldon produces an accurate mathematical theory of human behavior.

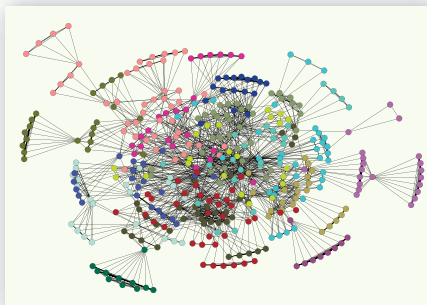
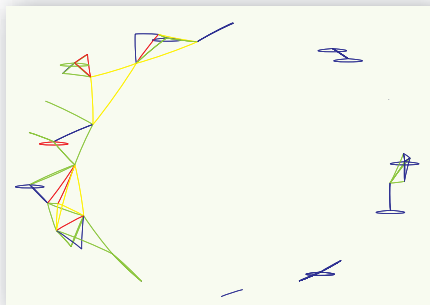
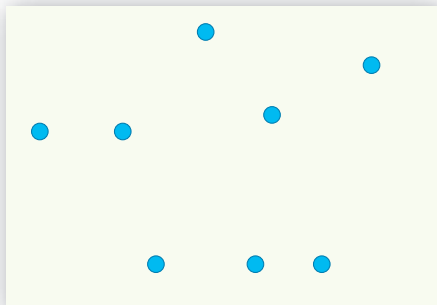
It may be hard to believe, but conditions for such a quantitative approach are increasingly in place. Indeed, records of human

actions are already stored in numerous databases. E-mail and phone records document our social and professional interactions; travel records and GPS navigation systems capture our travel patterns and physical locations; credit-card companies maintain records of our shopping and entertainment habits. Although in the wrong hands, these data sets represent Orwellian tools of power, for scientists they offer incredible insights into human behavior. Combine this capability with the sophisticated tool of network theory (2–7), which analyzes relations between millions of individuals, and you get a glimpse of an unprecedented opportunity to quantify human dynamics. Although a mathematical theory of social complexity remains a pipe dream, it is not as farfetched as it may have appeared in 1942, when *Foundation* was first published. Proof of this can be found in the study by Guimerà *et al.* on page 697 of this issue (8). By taking advantage of publicly

available data sets from both artistic and scientific fields, the authors offer powerful insights into the mechanisms governing collective human behavior.

Traditionally, the achievements of individuals such as Darwin and Einstein have dominated the public's image of science, yet today some of the most groundbreaking work is collaborative in nature (see the figure). But how do such creative teams come about? Are there discernible differences between collaborations that are sparkingly creative and those that are less inventive? Guimerà *et al.* use network theory to answer these questions. Their starting point is a collection of fascinating data sets: a century-long record of Broadway musicals and the publication records of several fields of science. These data sets allowed them to reconstruct the collaborative history of the individuals who contributed to a particular show or research publication. The investigators document a changing creative enterprise in which advances require an increasing number of contributors. The history of Broadway is particularly illuminating: The team size responsible for producing a show increased until the 1930s, after which it leveled off, fluctuating at around seven contributors for the past 70 years. In contrast, science continues to search for its optimal collaborative setup: The number of

The author is in the Center for Complex Network Research and the Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA. E-mail: alb@nd.edu



Evolution of the scientific enterprise. (Left) For centuries, creative individuals were embedded in an invisible college, that is, a community of scholars whose exchange of ideas represented the basis for scientific advances. Although intellectuals built on each other's work and communicated with each other, they published alone. Most great ideas were attributed to a few influential thinkers: Galileo, Newton, Darwin, and Einstein. Thus, the traditional scientific enterprise is best described by many isolated nodes (blue circles). **(Middle)** In the 20th century, science became an increasingly collaborative enterprise, resulting in such iconic pairs as the physicist Crick and the biologist Watson (left),

who were responsible for unraveling DNA's structure. The joint publications documenting these collaborations shed light on the invisible college, replacing the hidden links with published coauthorships. **(Right)** Although it is unlikely that large collaborations—such as the D0 team in particle physics or the International Human Genome Sequencing Consortium pictured here—will come to dominate science, most fields need such collaborations. Indeed, the size of collaborative teams is increasing, turning the scientific enterprise into a densely interconnected network whose evolution is driven by simple universal laws.

coauthors in each scientific field has increased monotonically during the past decade. It is anyone's guess when and where it will reach a maximum.

Until the late 1990s, the bulk of network research focused on static properties, which do not change with time (9). Yet a proper understanding of most networks requires that we characterize the assembly process that generated them. Indeed, a map of such networks is not sufficient to understand the structure of the World Wide Web—we must describe how documents and links are added and removed (3). Uncovering all interactions between proteins is only the first step toward understanding cellular networks—we must also explore the importance of gene duplications and mutations that shape the interactions between proteins and genes (10). Similarly, to comprehend the structure of the collaboration map, we must understand how people form friendships and alliances. Given that in the professional world friendships are just as crucial as hard-nosed professional interests, modeling the evolution of creative teams may appear to be impossible. Guimerà's results indicate otherwise: They show that a simple model successfully captures many qualitative features of the network underlying the creative enterprise. In their study, they distinguish between veterans, who have participated in collaborations before, and rookies, who are about to see

their names appear in print for the first time. Two parameters are key: the fraction of veteran members in a new team, and the degree to which veterans involve their former collaborators. If choosing experienced veterans is not a priority, the authors find that the network will be broken up into many small teams with little overlap between them. As the likelihood of relying on veterans increases, thanks to the extra links to earlier collaborators, the teams coalesce through a phase transition such that all players become part of a single cluster.

Many professional networks—from the web of actors in Hollywood to scientific collaborations (11, 12)—are scale-free (13), that is, although most individuals have only a few collaborators, a few have hundreds and operate as hubs. The legendary Paul Erdős, the father of random network theory, with more than 500 collaborators, was probably the best known hub within mathematics. The model that Guimerà and co-workers propose does indeed account for hubs, the emergence of which is rooted in the rookies' desire to involve their friends in new teams. Indeed, the more collaborators an individual has, the higher the chances are that he or she will be invited to participate again. This process—called preferential attachment in network theory—is responsible for the emergence of hubs through a rich-gets-richer process (13) in which well-connected individuals continue to be in high demand.

How does this assembly process affect the team's performance? The results of the Guimerà *et al.* study indicate that expertise does matter: Teams publishing in high-impact journals have a high fraction of incumbents. But diversity matters too: Teams with many former collaborative links offer inferior performance. Thus, the recipe for success seems relatively simple: When forming a "dream team" make an effort to include the most experienced people, whether or not you have worked with them before. The temptation to work mainly with friends will eventually hurt performance.

In Asimov's classic story, Harry Seldon's theory could not handle innovation. To stay on the predictive side, the Foundation went to great lengths to freeze all technological development. Indeed, the most disruptive social changes humanity has experienced are intimately tied to new technologies, from the steam engine to the Internet. It is tempting to conclude, therefore, that given the unpredictability of potential technologies, a theory of human dynamics will have no chance of success until scientific innovation ceases. A more constructive approach, and one taken by Guimerà *et al.*, takes us in the opposite direction, bringing innovation into a scientific and mathematical perspective.

Finally, will there ever be a Harry Seldon and a mathematical theory of human behavior? It is easy to maintain that human actions

PHOTO CREDITS: (LEFT) NEWTON AND GALILEO/ANP; EMILIO SEGRE VISUAL ARCHIVES; DARWIN/SOTHOBY'S; (MIDDLE) A. BARRINGTON BROWN/PHOTO RESEARCHERS; (RIGHT) LARRY THOMPSON/NIH

are too complex to be predictable. But skeptics are proven wrong each time a waiter brings them ketchup with their fries, without having been asked to. A master of consumer behavior, the waiter concludes that very likely they will ask for it. In the same way, a data-driven understanding of human actions could help us to translate into a predictive mathematical language the fundamental principles that drive a society's collective behavior. In a world in which all events are recorded by computers, the conditions for this research are increasingly in place. The quantitative accumulation of such data could easily spark a qualitative change, helping many disparate facts to fall into a coherent predictive theory. By demonstrating that the Web, the cell, or society is driven by similar organizing principles, network

theory offers a successful conceptual framework to approach the structure of many complex systems. Perhaps a predictive framework that captures the dynamics and the behavior of networks is not too far behind either. In the light of Guimerà *et al.*'s results, we can be sure of one thing: While pursuing a theory of human behavior, we could overlook a Harry Seldon. A mathematical theory of human dynamics may not be the solitary achievement of a genius scientist (14), but will likely emerge from the combined efforts of an expert team with just the right combination of expertise, collaborative experience, and fresh ideas.

References

1. I. Asimov, *Foundation and Empire* (Spectra, New York, 1991).
2. S. N. Dorogovtsev, J. F. F. Mendes, *Evolution of*

Networks: From Biological Nets to the Internet and WWW (Oxford Univ. Press, New York, 2003).

3. A. Pastor-Satorras, A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge Univ. Press, Cambridge, UK, 2004).
4. R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
5. S. H. Strogatz, *Nature* **410**, 268 (2001).
6. E. Ben-Naim, H. Frauenfelder, Z. Toroczkai, *Complex Networks* (Springer Verlag, Berlin, 2004).
7. S. Bornholdt, H. G. Schuster, Eds., *Handbook of Graphs and Networks: From the Genome to the Internet* (Wiley-VCH, Weinheim, Germany, 2003).
8. R. Guimerà, B. Uzzi, J. Spiro, L. A. Nunes Amaral, *Science* **308**, 697 (2005).
9. B. Bollobás, *Random Graphs* (Cambridge Univ. Press, Cambridge, UK, 2001).
10. A.-L. Barabási, Z. N. Oltvai, *Nat. Rev. Genet.* **5**, 101 (2004).
11. M. E. J. Newman, *Phys. Rev. E* **64**, 016131 (2001).
12. A.-L. Barabási *et al.*, *Physica A* **311**, 590 (2002).
13. A.-L. Barabási, R. Albert, *Science* **286**, 509 (1999).
14. R. Collins, *The Sociology of Philosophies* (Harvard Univ. Press, Cambridge, MA, 1998).

10.1126/science.1112554

OCEAN SCIENCE

Ocean Mixing in 10 Steps

Bill Merryfield

Ocean waters are warm in some places, cold in others, and also vary widely in their salt content, or salinity. As masses of water transit the globe in ocean currents, these properties are modified by air-sea exchanges (including warming by sunshine and freshening by rain) and by subsurface hydrodynamic processes referred to as ocean mixing (1). On page 685 of this issue, Schmitt *et al.* (2) report direct measurements of one such mixing process.

Measurements of this kind are important because the temperature and salinity of a water mass govern its buoyancy, and hence determine how it rises or sinks across ocean surfaces of constant density. For example, the Atlantic overturning circulation, which transports heat from tropical to subpolar regions, is supplied with sinking water through buoyancy loss (mainly from surface cooling) in the far North Atlantic. For the water mass to complete the circuit, the lost buoyancy must be regained further south through some combination of air-sea exchange and ocean mixing.

Knowledge of ocean mixing is thus a prerequisite for understanding ocean circulation. Such understanding is greatly aided by ocean circulation models. Although these models have become more sophisti-

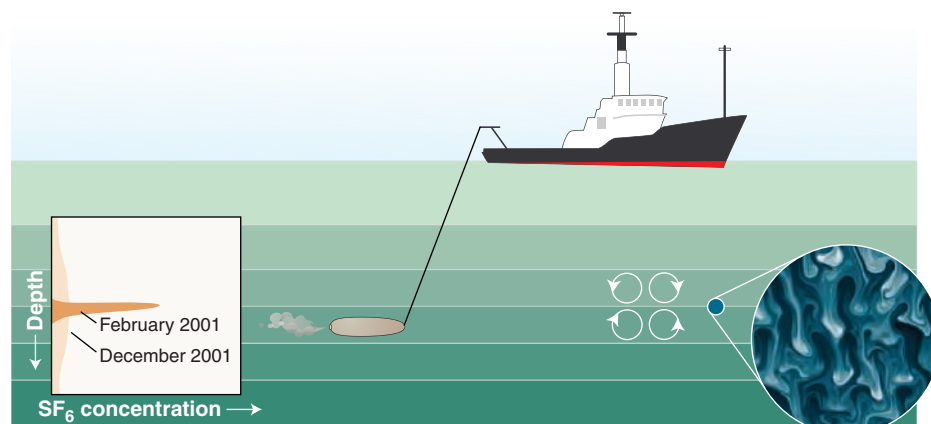
cated in recent years (3), much room for progress remains in how they treat mixing, which occurs on spatial scales much smaller than the models can represent explicitly. To this end, measurements like those of Schmitt *et al.* (2) provide vital guidance.

In the absence of extensive data, early models imposed relatively strong mixing that was either uniform or a prescribed function of depth. The prescribed values were consistent with theoretical estimates of mixing that likewise assumed horizontal uniformity (4). Meanwhile, indirect evidence was accumulating that mixing throughout much (and perhaps most) of the ocean might actually be much weaker. However, these meas-

urements relied on theoretical constructs to link small-scale temperature and velocity fluctuations to mixing (5, 6), and were thus not immune from skepticism.

A more definitive answer was provided by revolutionary direct measurements that involved injection of a nearly inert compound, sulfur hexafluoride (SF_6), into the ocean. SF_6 can be detected in minute concentrations months or even years after injection. Three large-scale experiments of this kind have been performed to date. The first of these, the North Atlantic Tracer Release Experiment of 1992 to 1994, followed the vertical spread of SF_6 about the 300-m injection depth and showed that conclusions drawn from earlier indirect measurements were substantially correct (7).

Though uniquely definitive, tracer-release experiments require large commitments of funding and ship time and therefore must be carefully targeted. The second study, from 1996 to 1998, involved the



Staircase mixing in the ocean. In early 2001, 175 kg of inert SF_6 were released into a "thermohaline staircase" in the western tropical Atlantic. Subsequent vertical dispersion of this tracer (**inset, bottom left**), measured 10 months later, revealed the extent of mixing by salt fingers in the thin interfaces (**inset, bottom right**) and by convection within the thicker layers. The mixing rate, which applies to salinity, was approximately double that of heat.

The author is with the Canadian Centre for Climate Modelling and Analysis, Meteorological Service of Canada, Victoria, British Columbia, V8W 2Y2 Canada. E-mail: bill.merryfield@ec.gc.ca

release of SF₆ at a depth of 4000 m in the Brazil Basin of the South Atlantic. It revealed a very different regime, where mixing is 10 to 100 times as strong as at the lesser depths sampled by the 1992 to 1994 experiment (8). Indirect measurements spanning the basin were also performed that detected elevated mixing in the deepest ocean over jagged features such as mid-ocean ridges, but not over smooth abyssal plains (9). The findings sparked a flurry of studies that converged on a common conclusion: Tidal currents flowing over peaks and valleys launch undersea waves, which in turn power elevated mixing (10, 11). This process largely governs the structure of the abyssal ocean and contributes substantially to the dissipation of tides.

The third major tracer-release experiment is described by Schmitt *et al.* (see the figure) (2). Performed in 2001, it targeted yet another distinct mixing regime which, though modest in extent, is exotic. In the upper tropical Atlantic, immediately east of the Caribbean, the ocean organizes itself into 10 or so steps, each 10 to 30 m high, which can be laterally coherent for hundreds of kilometers. Each step consists of a well-mixed layer bounded above and below by rel-

atively thin interfaces in which temperature and salinity decrease sharply with depth.

This grand “thermohaline staircase” clearly owes its existence to salt fingers, which are centimeter-scale rising and sinking tendrils that develop where warmer and saltier water overlies cooler, fresher water. In laboratory and numerical experiments, salt fingers mix salinity much more effectively than they mix heat (12, 13), but such a dichotomy had not been firmly established in the ocean. The difference in mixing rates arises (as does salt fingering itself) because molecular diffusion of salt is 100 times slower than the diffusion of heat; thus, excess salinity is “locked” into sinking fingers, whereas excess heat tends to leak away. Because SF₆ diffuses at nearly the same rate as salt, its transport by salt fingering should also be comparable.

Based on SF₆ dispersion, Schmitt *et al.* deduce a mixing rate for salinity that is approximately double that inferred indirectly for temperature. This rate is intermediate between the mixing rates in the upper eastern North Atlantic (7) and those in the deep Brazil Basin (8). A crucial property of mixing by salt fingering is that it tends to enhance the density contrasts between the abyssal and

upper ocean, whereas mixing driven by tides and winds reduces the density contrasts.

The results reported by Schmitt *et al.* (2) supply one more piece of a puzzle that oceanographers must assemble to achieve a fuller understanding of ocean mixing. This understanding will in turn lead to more accurate and reliable models of the oceans and of climate.

References and Notes

1. Lateral stirring by ocean eddies, leading ultimately to mixing through pathways that are not well understood, also plays a role. Land-sea exchanges such as geothermal heating and river runoff are important in certain locations.
2. R. W. Schmitt, J. R. Ledwell, E. T. Montgomery, K. L. Polzin, J. M. Toole, *Science* **308**, 685 (2005).
3. S. M. Griffies *et al.*, *Ocean Model.* **2**, 123 (2000).
4. W. Munk, *Deep-Sea Res.* **13**, 707 (1966).
5. T. R. Osborn, C. Cox, *Geophys. Fluid Dyn.* **3**, 321 (1972).
6. T. R. Osborn, *J. Phys. Oceanogr.* **10**, 83 (1980).
7. J. R. Ledwell, A. J. Watson, C. S. Law, *J. Geophys. Res.* **103**, 21499 (1998).
8. J. R. Ledwell *et al.*, *Nature* **403**, 179 (2000).
9. K. L. Polzin, J. M. Toole, J. R. Ledwell, R. W. Schmitt, *Science* **276**, 93 (1997).
10. C. Garrett, *Science* **301**, 1858 (2003).
11. E. Kunze, S. G. Llewellyn Smith, *Oceanography* **17**, 55 (2004).
12. R. W. Schmitt, *Prog. Oceanogr.* **56**, 419 (2003).
13. J. Yoshida, H. Nagashima, *Prog. Oceanogr.* **56**, 435 (2003).

10.1126/science.1111417

STRUCTURAL BIOLOGY

Nature's Rotary Electromotors

Wolfgang Junge and Nathan Nelson

Molecular motors abound in the cell. Myosin motors power muscle contraction, kinesin motors move vesicles from one end of the cell to the other, and the ribosome processes along RNA. These linearly operating molecular motors are all powered by cleavage of the universal “fuel” molecule ATP (adenosine 5′-triphosphate). The ATP synthase (or F-ATPase), which produces ATP, is a fine example of one of nature's rotary motors. F-ATPase consists of two coupled motors, one electrically driven and the other chemically driven. There are several types of rotary motors, but only three are electrically driven: the F₀-portion of the F-type ATPase, the V₀-portion of V-type ATPases, and the flagellar motor of bacteria. The first two obey similar construction principles, whereas the bacterial flagellar motor is quite different. But all three types of rotary motor contain a central, ion-binding rotor ring that is embedded in the respective cou-

pling membrane of the cell. The first high-resolution crystal structures of this ring are now revealed by Meier *et al.* on page 659 (1) and Murata *et al.* on page 654 (2) of this issue. Meier and colleagues report the structure at 0.24-nm resolution of the c ring of the F-type Na⁺-ATPase from *Ilyobacter tartaricus* (1). Meanwhile, Murata and co-workers present the structure at 0.21-nm resolution of the K ring of the V-type Na⁺-ATPase from the bacterium *Enterococcus hirae* (2). Some of the newly revealed properties are in keeping with current mechanistic models, whereas others defy previous postulates.

F-type ATPases usually synthesize ATP at the expense of ion-motive force, whereas V-type ATPases generate ion-motive force at the expense of ATP hydrolysis. Some ATPase subunits share sequence homology, whereas others are unique to each ATPase family (3). ATPases transport either protons (H⁺) or less commonly sodium cations (Na⁺) across their respective coupling membrane. These enzymes are constructed from two rotary motors—the ion-driven motor F₀ and the chemical generator F₁ in F-type ATPases, and the V₀ and V₁ motor/generators in V-type ATPases. The membrane-embedded F₀ and

V₀ domains mediate the movement of either protons or Na⁺ ions across the membrane, and the peripheral F₁ and V₁ domains interact with ADP, inorganic phosphate, and ATP (see the figure). F₀ and F₁, and V₀ and V₁, respectively, are mechanically coupled by a central rotating shaft and are held together by an eccentric stalk. The central shaft together with the ring to which it is firmly attached is termed the rotor, and the rest of the ATPase is called the stator. The ion-driven rotation of the central c ring in F-type ATPases or K ring in V-type ATPases relative to the peripheral subunits of the F₀ or V₀ domains generates torque. Transmission of this torque to the F₁ or V₁ domains drives them to operate over the three C₃-symmetrical reaction sites where ATP is assembled from or cleaved into ADP and inorganic phosphate in a stepped rotation. The c ring in F₀ is composed of 10, 11, or 14 identical polypeptides depending on the organism (4–6). The 10 to 14 steps that mark progression of the c ring are adapted to the nonmatching C₃ symmetry of F₁ owing to the elasticity of the rotary power transmission between the two motors, which is essential for their working together efficiently (7, 8). Previous structural data for the c ring of F₀ were derived from nuclear magnetic resonance (NMR) (9), cryo-electron microscopy (cryo-EM) (10), x-ray crystallography (6), and atomic force microscopy (4, 5). However, these structural data lacked sufficient resolution. An NMR

W. Junge is in the Division of Biophysics, University of Osnabrück, 49069 Osnabrück, Germany. E-mail: junge@uos.de N. N. Nelson is in the Department of Biochemistry, Tel Aviv University, 69978 Tel Aviv, Israel. E-mail: nelson@post.tau.ac.il

structure of the monomeric c subunit in a mixed solvent (9, 11) has been widely used in modeling studies, and also for refining one x-ray-derived model of the c ring (6).

The new crystal structures of both the c ring of F_0 (1) and the K ring of V_0 (2) reveal a concave barrel with a pronounced waist in the middle, and an inner septum that is probably filled with and electrically sealed by membrane lipids in vivo. The K ring of V_0 is about twice as wide as the c ring of F_0 , and Na^+ binding sites are found in the waist of both F_0 (1) and V_0 (2). Each protomer in the ring carries an essential, ion-binding glutamic acid

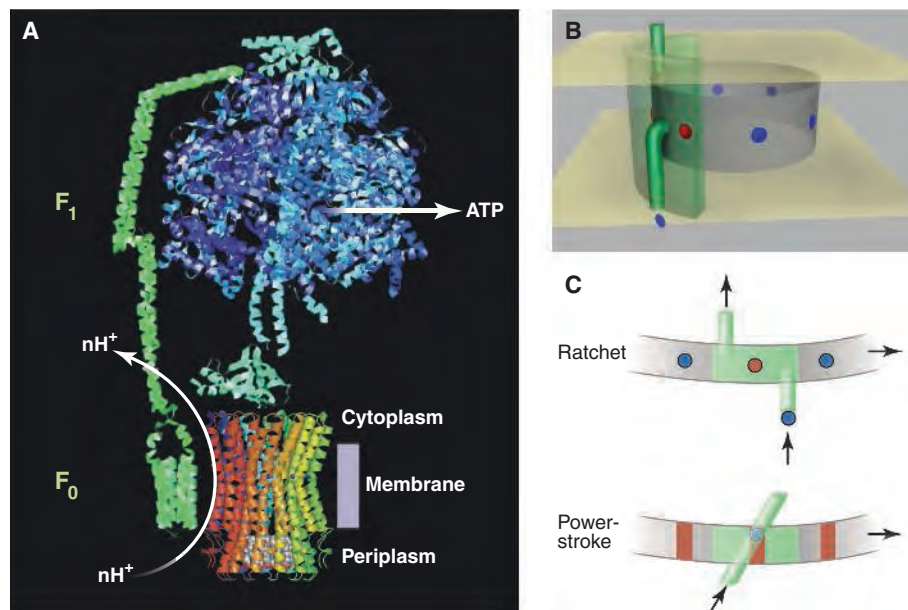
early located), thus creating chirality—complete the electromotor. A cation that enters through the lower access channel and binds to the essential glutamic acid residue in the ring relieves the electrostatic constraint such that the ring is allowed to move one step farther counterclockwise. Thus, the cation concentration gradient determines the direction of rotary motion: counterclockwise if it is positive in the lower compartment, and clockwise if it is negative in the lower compartment. In this way, ion flow generates torque. This Brownian ratchet generates directed motion based on stochastic thermal fluctuations of the

(13, 17) is not supported by the new ring structures, as Murata *et al.* point out (2). Neither the c-ring nor K-ring structures provide any obvious clues about the free accessibility from one side of the essential glutamic acid residue in the ring, but note the argument made by Meier *et al.* (1), who still favor free accessibility from one side.

The two-channel-and-ring Brownian ratchet (12) remains a viable model for torque generation in both the F-type and V-type electromotors. An alternative is the power-stroke model (see the figure, C). In this model, a cation moves across the membrane only when positioned between two half-cylindrical channels at the point of their overlap, one half-channel being on the rotor and the other on the stator (18). Because of the oblique orientation of one half-channel and the straight orientation of the other, a moving cation forces the ring to rotate. This simple power-stroke-only mechanism (see the figure) is difficult to accept because the new ring structures show the bound ions moving around with the ring. The Brownian ratchet thus remains the most probable mechanism. Whether there are partial power-stroke elements at the interface of contact between the ring and the stator might be revealed by future high-resolution structures of the stator elements apposed to the ring.

It has come as a surprise to those working on V-type ATPases that the crystal structure of the Na^+ -translocating K ring shows decameric symmetry (2). Indeed, Murata and co-workers recently proposed a heptameric structure on the basis of single-particle classification of cryo-EM images (19). On the one hand, this discrepancy will certainly prompt a reconsideration of the reliability of single-particle image analyses. On the other hand, it raises the question of whether the present K-ring structure of the V-type ATPase from *E. hirae* can be extrapolated to the V-type ATPases of other organisms.

The structural symmetry of the F_0 c ring (10–14) versus the c ring of F_1 implies a H^+ -to-ATP stoichiometry of 3.3 to 4.7 (which favors ATP formation). In eukaryotic V-type ATPases, six (double hairpin-shaped) c subunits, each with only one ion-binding residue, may compose the c ring, equivalent to the K ring in *E. hirae*. This suggests an H^+ -to-ATP stoichiometry of 2 to 1 (6 to 3), which would favor ATP hydrolysis and proton pumping against a large pH difference by the V-ATPase (20). This low gear ratio has been held responsible for the generation of a very acidic internal pH in certain organelles, for example, pH 2 in vacuoles of lemon fruit cells. This acidic pH, as well as a voltage of 250 mV generated by a V-type ATPase in the insect midgut, is difficult to reconcile with the presence of 10 K subunits forming the rotor ring of the V-type ATPases [gear ratio of 3.3 (10 to 3)]. The sug-



A turning point for ATPase. (A) The F-type ATPase is composed of the F_1 domain—which interacts with ATP, ADP, and phosphate—and the F_0 domain, which moves ions across the respective coupling membrane. Both domains are mechanically coupled by a central rotary shaft and an eccentric stalk. The membrane-embedded F_0 domain contains the c ring, formed from identical hairpin-shaped protomers. Ion flow (here proton flow) spins the c ring relative to peripheral subunits of F_0 , thereby generating torque, which is transmitted by the rotary shaft to the F_1 domain. F_1 carries three reaction sites for ATP synthesis/hydrolysis. (The figure is composed of several partial structures of F-ATPase from various organisms.) (B) The Brownian rotary ratchet mechanism of ion conduction by an F-type ATPase (12, 27). (C) Ratchet versus power-stroke mechanism of torque-generating ion translocation.

residue (Glu) in the middle of the respective carboxyl-terminal helix: Glu⁶⁵ in the hairpin-shaped subunit c of F_0 from *I. tartaricus* (1) and Glu¹³⁹ in the subunit K double-hairpin of V_0 from *E. hirae* (2). To appreciate the new features of both structures, we need to consider how F_0 and V_0 may generate torque (12–14). The ring is thought to undergo Brownian rotational fluctuations relative to the membrane portion of the stator (see the figure). Fluctuations are subject to two electrostatic constraints: The essential glutamic acid residues in the ring must be negatively charged when facing the positive stator and neutralized by cation binding when facing hydrophobic membrane lipids (see the figure). Two access channels for ions—that are on either side of the stator, (and are not colin-

ring against the stator and of ions flowing into and out of the access channels. Since it was first proposed, this concept has been modified in two different ways: by invoking a swiveling motion of one of the two helices of the hairpin, and by assuming that the essential glutamic acid residue in the repetitive hairpins is directly accessible from one aqueous compartment. Both of these modifications are not supported by the new structures. First, the new c-ring crystal structure of F_0 , unlike the previously assumed one, is now fully compatible with c-to-a cross-linking data (15, 16). This eliminates the need to assume swiveling as pointed out by Meier and colleagues (1). Instead, the electromotor might operate by rotation of the rigid ring. Second, the one-channel model proposed by Dimroth's group

gestion by Murata *et al.* (2) that the same ratio may also be true for eukaryotic V-type ATPases implies that we may have to rethink how an internal pH of 2 is formed in the vacuoles of lemon cells or just accept that the c ring of eukaryotic V-type ATPases (assembled from three different gene products) has a different symmetry. High-resolution structures of the V_0 domain of the eukaryotic V-type ATPase will hopefully resolve this issue.

References and Notes

1. T. Meier, P. Polzer, K. Diederichs, W. Welte, P. Dimroth, *Science* **308**, 659 (2005).

2. T. Murata, I. Yamato, Y. Kakinuma, A. G. W. Leslie, J. E. Walker, *Science* **308**, 654 (2005); published online 31 March 2005 (10.1126/science.1110064).

3. N. Nelson, W. R. Harvey, *Physiol. Rev.* **79**, 361, (1999).

4. H. Stahlberg *et al.*, *EMBO Rep.* **2**, 229 (2001).

5. H. Seelert *et al.*, *Nature* **405**, 418 (2000).

6. D. Stock, A. G. Leslie, J. E. Walker, *Science* **286**, 1700 (1999).

7. W. Junge *et al.*, *FEBS Lett.* **504**, 152 (2001).

8. O. Pänke, D. A. Cherepanov, K. Gumbiowski, S. Engelbrecht, W. Junge, *Biophys. J.* **81**, 1220 (2001).

9. M. E. Girvin *et al.*, *Biochemistry* **37**, 8817 (1998).

10. J. Vonck *et al.*, *J. Mol. Biol.* **321**, 307 (2002).

11. V. K. Rastogi, M. E. Girvin, *Nature* **402**, 263 (1999).

12. W. Junge, H. Lill, S. Engelbrecht, *Trends Biochem. Sci.* **22**, 420 (1997).

13. P. Dimroth, H. Wang, M. Grabe, G. Oster, *Proc. Natl. Acad. Sci. USA* **96**, 4924 (1999).

14. W. Junge, *Photosynth. Res.* **80**, 197 (2004).

15. R. H. Fillingame, W. Jiang, O. Y. Dmitriev, P. C. Jones, *Biochim. Biophys. Acta* **1458**, 387 (2000).

16. R. H. Fillingame, O. Y. Dmitriev, *Biochim. Biophys. Acta* **1565**, 232 (2002).

17. P. Dimroth, U. Matthey, G. Kaim, *Biochim. Biophys. Acta* **1459**, 506 (2000).

18. P. Lauger, *Biophys. J.* **53**, 53 (1988).

19. T. Murata *et al.*, *J. Biol. Chem.* **278**, 21162 (2003).

20. H. Arai, G. Terres, S. Pink, M. Forgac, *J. Biol. Chem.* **263**, 8796 (1988).

21. For an animation see www.biologie.uni-osnabrueck.de/biophysik/junge.

10.1126/science.1112617

NEUROSCIENCE

Understanding Intentions: Through the Looking Glass

Kiyoshi Nakahara and Yasushi Miyashita

If you see a person sitting at a dining table and scooping out soup with a spoon, you naturally understand the person's intention is to eat the soup with the spoon. However, if a baby is sitting next to that person, the action of scooping out the soup may be followed by the action of feeding the baby. When we understand the intentions of other people through their actions, "action" (in this case, scooping out the soup) should be understood in context. What brain mechanisms link action understanding to a particular context? With regard to action understanding, a class of neurons in the brain called "mirror neurons" are thought to be important because of their amazing ability to transmit electrical impulses not only during the execution of a particular action, but also during observation of the equivalent action being carried out by someone else (1–4). The name mirror neuron reflects this remarkable degree of visuo-motor congruency. If mirror neurons also are involved in the understanding of intention, then they must be activated according to the context in which the action occurs. On page 662 of this issue, Fogassi *et al.* (5) explore mirror neurons in the inferior parietal cortex of the monkey brain, and report a relationship between mirror neurons and the encoding of the context under which actions occur. Their findings suggest that mirror neurons in the inferior parietal cortex may be responsible for understanding the intention of action in other people.

Mirror neurons were originally discovered in monkeys, first in the ventral premotor

cortex of the frontal lobe and then in the inferior parietal cortex (2, 6). Recordings from single neurons in the monkey brain revealed several interesting features of mirror neurons in addition to their congruent visuomotor responses (2). First, activation of mirror neurons appears to require biological motion; that is, even if a mirror neuron shows a specific response to the observation of grasping, it does not respond when grasping is achieved using an artificial tool such as pliers. Second, not only visual information but also the sound of a particular action (such as tearing paper) can induce mirror neurons to fire. Third, activation of mirror neurons requires the achievement of the goal of the action. For example, the observation of grasping an object can activate mirror neurons, but the observation of fake grasping in the absence of an object, or of the object but without any grasping action, cannot evoke such discharges. Interestingly, under conditions where monkeys clearly understand that the goal of action has been achieved, mirror neurons respond even when the end of the action is occluded from their sight.

The third property, the requirement of goal-directed action, prompts an expectation that, to some degree, mirror neurons participate in understanding the intention underlying action. However, because previous experiments identified this goal dependency only in terms of the contact between individuals and target objects, this property alone seems to be insufficient for understanding intentions that require information about the context or the "hypergoal" of actions. In their new work, Fogassi *et al.* (5) recorded the discharges of neurons in the rostral part of the inferior parietal cortex

(areas PF and PFG) of the monkey brain as the monkeys carried out a behavioral test with different action contexts. These investigators identified neurons that not only have mirror properties but also are able to encode actions in a context-dependent way.

The authors introduced two action contexts (see the figure, B). Under one set of circumstances, monkeys reach toward and grasp a food pellet, then eat it; under the other set, monkeys reach toward and grasp a non-food pellet and then place it inside a container. The action of grasping an object is a component common to the two situations, but one that is followed by two different final goals: grasp-to-eat or grasp-to-place. One population of neurons showed discharges specifically related to the grasping action. The activity of these grasping-related neurons was often modulated by the context of the action. For example, some of these neurons showed higher discharge rates during grasping followed by eating relative to grasping followed by placing, and some showed the opposite behavior. Control experiments supported the idea that this modulation was specifically dependent on the circumstance of the action, and was not due to differences in the target object (food or nonfood) or to kinematics (the path or velocity of motion.) These findings suggest that neurons in the inferior parietal cortex encode particular actions in context-dependent ways.

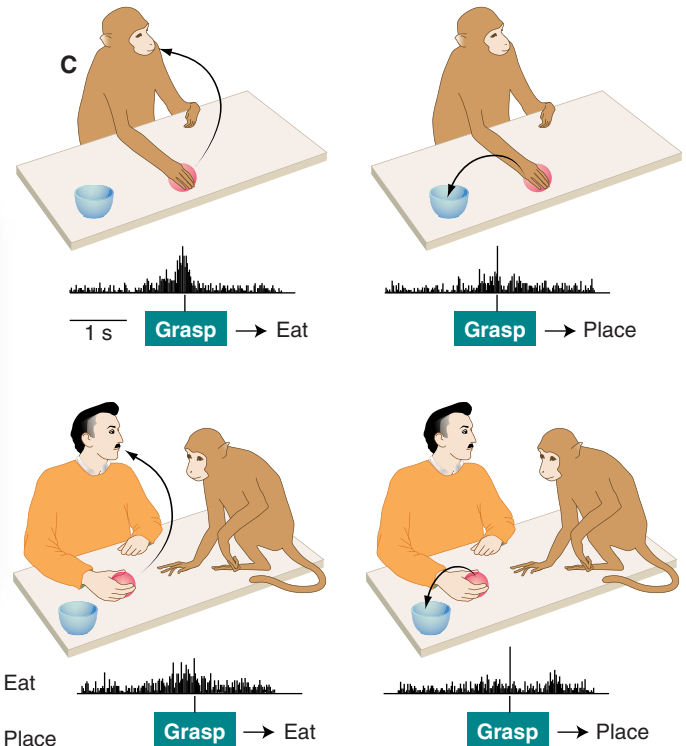
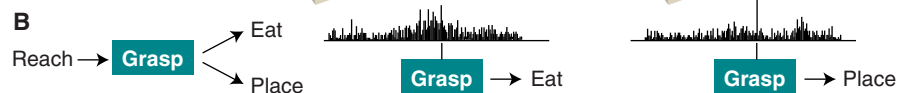
Of even more interest, the authors reveal that this context-dependent encoding of action has a close relationship with the mirror properties of neurons in areas PF and PFG (see the figure, C). They examined mirror neurons that fire during both grasping execution and grasping observation. Among these, some mirror neurons also exhibited context-dependent modulation, and selectivity of this modulation was preserved across both execution of the action and observation of the action carried out by another individual. Thus, if a mirror neuron exhibited higher firing rates during execution of a grasp-to-eat action relative to a grasp-to-place action, this neuron also exhibited greater activation dur-

The authors are in the Department of Physiology, University of Tokyo School of Medicine, Bunkyo-ku, Tokyo 113-0033, Japan. E-mail: nakahara@m.u-tokyo.ac.jp; yasushi_miyashita@m.u-tokyo.ac.jp

ing observation of a grasp-to-eat action relative to a grasp-to-place action carried out by someone else.

Although the quantitative evidence is weak, the authors synthesize an attractive hypothesis from these results. They postulate that the coexistence of mirror properties and context-dependent action encoding in the inferior parietal cortex provides the basis for understanding intention (see the figure, A). Neurons in the intraparietal sulcus, located dorsal to the brain regions investigated in the Fogassi *et al.* study, have been implicated in several cognitive functions that require integration of sensory and motor information, such as movement planning, spatial attention, and decision-making (7–9). In particular, planning a specific movement can be considered a part of the representation of intention (9). On the other hand, the effect of attention can be a confounding factor for the situation-dependent modulation of mirror neuron activity presented here (7). Because grasp-to-eat seems to be a more intrinsic behavior for monkeys than grasp-to-place, it is possible that the grasp-to-eat task readily captures their attention. Indeed, the data presented here indicate that a greater number of inferior parietal neurons preferred a grasp-to-eat condition over a grasp-to-place condition. Regarding Fogassi *et al.*'s experimental design, it might have been better if different behavioral conditions had been balanced from an ecological point of view.

The imaginative experiments presented here raise further questions. First, so far, no experimental data exist concerning whether firing of mirror neurons in the ventral premotor cortex also show modulation by the action circumstances as is the case with mirror neurons in the inferior parietal cortex. Exploring functional differences between ventral premotor and inferior parietal mirror neurons is of particular interest. Second, cognitive deficits caused by the disruption of mirror neurons have not yet been clearly identified. Although some clinical case studies have suggested that damage around the frontal opercular region (which probably contains mirror neurons) causes deficits in understanding actions, experimental lesions in monkey mirror neurons will provide more direct and informative evidence of this. Third, as current hypotheses claim, the properties of mirror neurons may explain the



Through the mirror neuron: And what Alice understood there. (A) Mirror neurons can be activated not only when Alice reaches for Humpty Dumpty's hand but also when she looks at Humpty reaching for her hand. Through the activity of her mirror neurons, Alice can understand the action and intention of Humpty. (B) In their study, Fogassi *et al.* (5) provide monkeys with two action contexts: reaching toward and grasping a food pellet that the animals eat, and reaching toward and grasping a nonfood pellet that the animals place inside a container. Note that the action of grasping is common to the two contexts. (C) Top: Context-dependent modulation of mirror neuron activity in monkey parietal cortex during the execution of an action. The mirror neuron shows higher firing rates around the time of grasping followed by eating (left) relative to grasping followed by placing (right). Neuronal firing rates are indicated by peristimulus time histograms. Bottom: Context-dependent modulation of mirror neuron firing during observation of the action. The same neuron as in the top panel is also activated during the monkey's observation of grasping by an experimenter. Again, activity of the mirror neuron is modulated by the particular action context; that is, activity is higher when grasping is followed by eating relative to grasping followed by placing an object. Note that the preference for context dependency is preserved across the top and bottom panels. [(C) is modified from figure 5 of (5)]

neuronal mechanisms that underlie the ability to associate self and others through actions, such as the understanding of action and intention, imitation, and shared attention. However, such cognitive abilities are further developed and behaviorally more relevant in humans than in monkeys. The question is whether such human cognition can be explained by an extension of the findings with mirror neurons or whether there are additional neuronal mechanisms that are specific to humans. In particular, there is a debate concerning the possibility that mirror neurons constitute a prototype neuronal mechanism underlying human theory of mind or language ability (2–4). To investigate evolutionary aspects of the function of the mirror neuron system, we need comparative functional studies between monkeys and humans (10). Although elucidation of their precise roles in cognition in monkeys and humans awaits further experimental and

theoretical investigation, the astonishing properties of mirror neurons point to a unique principle that our brains use to link self to nonself.

References

1. W. Prinz, *Eur. J. Cognit. Psychol.* **9**, 129 (1997).
2. G. Rizzolatti, L. Fogassi, V. Gallese, *Nat. Rev. Neurosci.* **2**, 661 (2001).
3. P. L. Jackson, J. Decety, *Curr. Opin. Neurobiol.* **14**, 259 (2004).
4. S. J. Blakemore, C. Frith, *Neuropsychologia* **43**, 260 (2005).
5. L. Fogassi *et al.*, *Science* **308**, 662 (2005).
6. V. Gallese, L. Fadiga, L. Fogassi, G. Rizzolatti, *Brain* **119**, 593 (1996).
7. C. L. Colby, M. E. Goldberg, *Annu. Rev. Neurosci.* **22**, 319 (1999).
8. J. I. Gold, M. N. Shadlen, *Trends Cognit. Sci.* **5**, 10 (2001).
9. R. A. Andersen, C. A. Buneo, *Annu. Rev. Neurosci.* **25**, 189 (2002).
10. K. Nakahara, T. Hayashi, S. Konishi, Y. Miyashita, *Science* **295**, 1532 (2002).

10.1126/science.1112174

The Influence of Social Hierarchy on Primate Health

Robert M. Sapolsky

Dominance hierarchies occur in numerous social species, and rank within them can greatly influence the quality of life of an animal. In this review, I consider how rank can also influence physiology and health. I first consider whether it is high- or low-ranking animals that are most stressed in a dominance hierarchy; this turns out to vary as a function of the social organization in different species and populations. I then review how the stressful characteristics of social rank have adverse adrenocortical, cardiovascular, reproductive, immunological, and neurobiological consequences. Finally, I consider how these findings apply to the human realm of health, disease, and socioeconomic status.

One of the greatest challenges in public health is to understand the “socioeconomic gradient.” This refers to the fact that in numerous Westernized societies, stepwise descent in socioeconomic status (SES) predicts increased risks of cardiovascular, respiratory, rheumatoid, and psychiatric diseases; low birth weight; infant mortality; and mortality from all causes (1–4). This relation is predominately due to the influence of SES on health, rather than the converse, and the disease incidences can be several times greater at the lower extreme of the SES spectrum.

One set of questions raised by the gradient concern its external causes. Despite human aversion to inequity in some settings (5), many Westernized societies tolerate marked SES gradients in health care access. Is this the predominant cause of the health gradient, or is it more a function of differences in lifestyle risk factors or of the psychosocial milieu in which poverty occurs?

Another set of questions concern the physiological mediators of the SES-health relationship—how, in a frequently used phrase in the field, does poverty get under the skin? These physiological questions are difficult to study in humans, and an extensive literature has focused instead on nonhuman animals. Despite the demonstration that some nonhuman species can also be averse to inequity (6), groups of social animals often form dominance hierarchies, producing marked inequalities in access to resources. In such cases, an animal’s dominance rank can dramatically influence the quality of its life. Does rank also influence the health of an animal?

The study of rank-health relations in animals has often been framed in the context of stress

and the idea that animals of different ranks experience different patterns of stress (Fig. 1). A physical stressor is an external challenge to homeostasis. A psychosocial stressor is the anticipation, justified or not, that a challenge to homeostasis looms. Psychosocial stressors typically engender feelings of lack of control and predictability and a sense of lacking outlets for the frustration caused by the stressor. Both types of stressor activate an array of endocrine and neural adaptations (Fig. 2). When mobilized in response to an acute physical challenge to homeostasis (such as fleeing a predator), the stress response is adaptive, mobilizing energy to exercising muscle, increasing cardiovascular tone to facilitate the delivery of such energy, and inhibiting unessential anabolism, such as growth, repair, digestion, and reproduction. Chronic activation of the stress response by chronic psychosocial stressors (such as constant close proximity to an anxiety-provoking member of one’s own species) can increase the risk of numerous diseases or exacerbate such pre-existing diseases as hypertension, atherosclerosis, insulin-resistant diabetes, immune suppression, reproductive impairments, and affective disorders (7).

In most social species, dominance rank influences the extent to which an individual sustains physical and psychosocial stressors. Thus, dominance rank can potentially influence an individual animal’s vulnerability to stress-related disease. In this review, I first consider which social ranks are most stressful, with an emphasis on nonhuman primates; stress can be experienced by both high- and low-ranking animals, and it varies as a function of the social organization in different species and populations. I then review the pathology that occurs in animals suffering from the most rank-related social stress. Finally, I consider the relevance of these hierarchy/health relationships to humans.

Which Ranks Are More Stressful?

No consensus exists as to whether dominant or subordinate animals are more physiologically “stressed.” Research in the 1950s, since discredited, argued that high rank was more physiologically stressful (that is, the “executive stress syndrome,” which was purportedly valid for both humans and other primates) (8). By the 1960s, the prevailing view had become that lower dominance rank carries the greatest risk of stress-related disease (9). It has now become clear that this too is an incorrect generalization. The contemporary view reflects the heterogeneity that is the core of ethology: Rank means different things in different species and populations. Patterns that occur amid this heterogeneity help to resolve many inconsistencies in the data, showing that the rank that experiences the most physical and psychological stressors tends to display the most severe stress-related pathologies (Fig. 2).

Resource inequity. The extent to which resources are divided unequally among individuals varies as a function of the dominance style of different species. At one extreme are top-down “despotic” hierarchies in which resource access is skewed markedly and dominant positions are attained through aggression and intimidation. In contrast, bottom-up “egalitarian” hierarchies have more equal resource distribution, and dominance is attained with the support of subordinate individuals (10). As will be seen, social subordination in despotic species can be associated with the greatest physiological indices of stress. In contrast, this is not a feature of subordination in egalitarian species.

Maintenance of dominance. In some species, rank is lifelong and inherited (for example, in female rhesus monkeys); in others, it may fluctuate, reflecting what has been aptly termed shifts in group “politics” (11). In species where ranks shift, how does an individual, once attaining a high rank, maintain it? At one extreme among species with despotic hierarchies, high-ranking individuals frequently and aggressively reassert their domination over the subordinate cohort (even in the absence of an overt challenge). In such species, which include dwarf mongooses, African wild dogs, and ring-tailed lemurs, dominant individuals have the greatest physiological indices of stress, most plausibly reflecting the physical demands of frequent fighting (12, 13). In contrast, in other

Departments of Biological Sciences, Neurology and Neurological Sciences, Stanford University, MC 5020, Stanford, CA 94305–5020, USA, and Institute of Primate Research, National Museums of Kenya. E-mail: sapolsky@stanford.edu

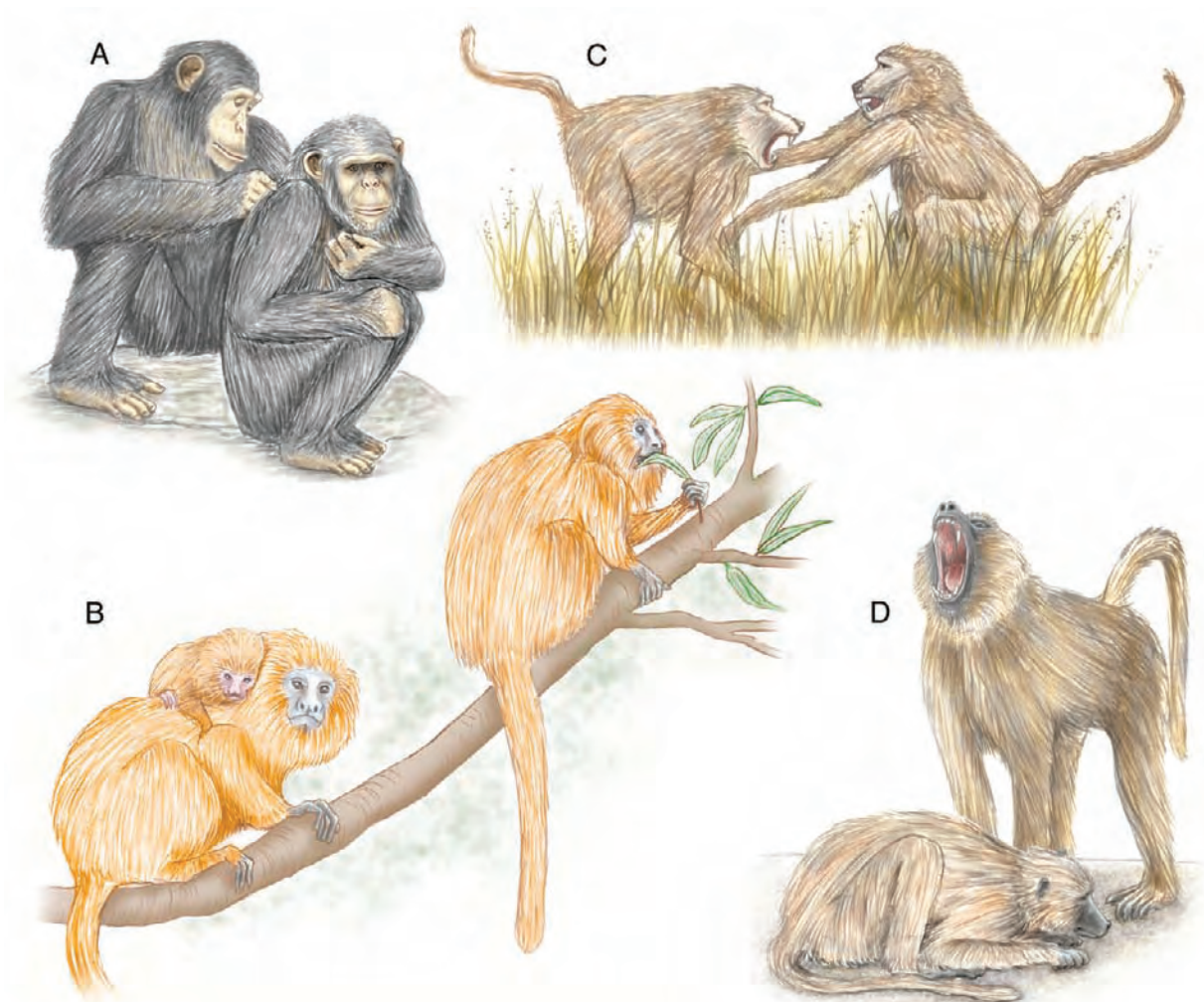


Fig. 1. (A and B) Affiliative behavior among subordinates can reduce the effects of stress. (A) Chimpanzees engage in social grooming. (B) A female tamarin monkey cares for another's young while the mother feeds. (C and D) Stressful dominance behavior may take physical or psychosocial forms. (C) Male savanna baboons may fight over a kill. (D) A dominant male baboon intimidates a subordinate. [Image credit: Carin Cain/Science]

despotic species, high-ranking individuals maintain dominance through psychological intimidation rather than aggression (where, for example, mere eye contact with the alpha individual might elicit subordination gestures). In such cases (e.g., savanna baboons, rhesus and squirrel monkeys, mice, rats, and white-throated sparrows), subordination is associated with the greatest physiological indices, plausibly reflecting the frequent psychological stressors for subordinates and the paucity of physical stressors for dominant individuals (12–18).

Breeding style. In many species, including some Old World primates, dominant alpha individuals of both genders monopolize breeding through aggression and intimidation. This can be sufficiently stressful to impair fertility in subordinates, producing “social contraception.” A different picture occurs in cooperative breeders, where one breeding female dominates other females, who are anovulatory. However, this subordination is minimally stressful, not involving aggression or harassment by the dominant female. Instead,

the anovulatory individuals are mostly younger sisters, waiting their turn to breed and helping to raise nieces and nephews (19). Among cooperative breeders such as marmosets, ring-tailed lemurs, marmots, wolves, and Florida scrub jays, subordinates show no more stress-related pathophysiology than do dominant individuals and may even have fewer indices (13, 19–21).

Stability of social ranks. When the hierarchy is stable in species where dominant individuals actively subjugate subordinates, it is the latter who are most socially stressed; this can particularly be the case in the most extreme example of a stable hierarchy, namely, one in which rank is hereditary. This reflects the high rates of physical and psychological harassment of subordinates, their relative lack of social control and predictability, their need to work harder to obtain food, and their lack of social outlets such as grooming or displacing aggression onto someone more subordinate. During major hierarchical reorganization, however, dominant individuals at the center of the social tensions typically experience the greatest amounts of physical and

psychological stress. As a result, during such reorganization among wild baboons or soon after group formation among species of captive primates, dominant individuals have the greatest physiological indices of stress; this has been shown in talapoin monkeys, squirrel monkeys, various macaque species, wild baboons, and chimpanzees. Once hierarchies stabilize, subordination becomes associated with the greatest physiological indices of stress (22).

Subordinate coping strategies. Stress-related physiological endpoints not only reflect the frequency and severity of stressors but also the availability and efficacy of coping outlets. Such outlets most commonly involve social support (such as grooming, physical contact, or coalition formation). Moreover, the occurrence in some species of reconciliative behaviors between two individuals shortly after a competitive interaction can be interpreted as a coping outlet for the loser of that interaction (23). The issue of coping outlets has been examined in a meta-analysis of rank-physiology relationships in both genders of an array of primate species.

Numerous variables related to social structure were considered, and three were collectively highly predictive of the occurrence of elevated stress hormone levels among subordinate animals: (i) high rates of being subjected to stressors; (ii) low availability of social support; and (iii) minimal presence of kin (24).

Subordinate avoidance of dominants. The inability to physically avoid dominant individuals is associated with stress, and the ease of avoidance varies by ecosystem. The spatial constraints of a two-dimensional terrestrial habitat differ from those of a three-dimensional arboreal or aquatic setting, and living in an open grassland differs from living in a plain dense with bushes. As an extreme example, subordinate animals in captivity have many fewer means to evade dominant individuals than they would in a natural setting (25). Thus, although dominant wolves have elevated stress hormone levels in the wild (21), subordinates demonstrate this trait in captivity (26).

Subordinates' use of alternative strategies. Implicit in being subordinate are the notions that one has reduced access to desirable resources and that this can translate into reduced Darwinian fitness. Sometimes, however, subordinate animals can pursue alternative behavioral strategies that, in effect, move them outside the hierarchy. For example, low rank among males of various Old World monkey species, as the result of male-male competition, has been thought to mean minimal reproductive access to females. However, females actually have considerable control over who they mate with. These are often low-ranking individuals with whom they have affiliative relationships (such as

frequent, nonsexual bouts of reciprocal grooming) (27). Such males not only have greater reproductive success than originally thought but also fewer physiological indices of stress than would be expected for their rank (28).

A different alternative strategy occurs among orangutans. Dominant males have pronounced secondary sexual characteristics, whereas subordinate individuals appear “juvenile.” This appearance is not merely a chronological stage. Instead, it is a state of arrested development in the presence of a dominant male and can persist for years. When the dominant male is removed, the apparently juvenile individual develops secondary sexual traits. This arrested state might seem to be a case of stress-induced social contraception. However, “juvenile” males are fertile, have some reproductive success (as they will force copulations when a dominant male is absent), and do not have elevated stress hormone levels or stress-related reproductive impairments. Rather than a stress-induced pathology, the arrest appears to be an alternative strategy. It is actually males in the process of the conspicuous, slow transition to the dominant form with the most marked physiological indices of stress (29).

Stress of dominating mating. In species with a sharply demarcated mating season, or where a few males disproportionately dominate mating, male-male competition for mating access can be fierce, dangerous, and at the cost of feeding and of affiliative behaviors. This raises the ironic possibility that dominant males may be sufficiently stressed by such competition that their testicular axes are suppressed. However, various endocrine mechanisms have evolved

that buffer reproductive physiology under that circumstance, either through blunting the release of stress hormones or blunting their ability to suppress the testicular system (30).

Atmosphere and culture. The nature of dominance varies with species and gender. Additionally, different populations of a species vary in their social milieu, and rank-physiology relationships can vary as well. For example, patterns of foraging by subordinate female spotted hyenas differ markedly between the enclosed Ngorongoro Crater and the open Serengeti Plains in East Africa, and only in the latter is subordination associated with elevated stress hormone levels (31). As another example, the elevated stress hormone levels observed among subordinate female macaques do not occur in a troop with atypically high rates of affiliative support (32, 33). In the realm of animal “culture,” multigenerational transmission of a culture of low aggression and high affiliation in a troop of wild baboons results in subordinate males that do not display the stress-related pathophysiology found in other troops (34).

Personality. Precedent exists for modulation of stress reactions by individuals' personalities. For example, independent of rank, primates who distinguish poorly between threatening and neutral stimuli, lack social outlets for support, and are hyperreactive to novelty have elevated stress hormone levels (35, 36) and increased rates of atherosclerosis (37).

Thus, under a variety of circumstances, social dominance can be associated with the most stress-related pathology, whereas in other situations, this is a trait of subordinate individuals. Are there common themes underlying this variability? Broadly and logically, adverse physiological profiles are most pronounced among animals of the rank exposed to the most physical and psychological stressors. This can arise from (i) low degrees of social control and predictability (as in dominant animals in unstable hierarchies and subordinate animals in small living spaces); (ii) a paucity of outlets after exposure to stressors (such as subordinate individuals in species lacking alternative strategies to hierarchical competition); (iii) a paucity of social support (for example, subordinate animals in settings with few kin and little access to social grooming); or (iv) high rates of physical stressors (such as dominant individuals who, as a function of their species or the instability of their hierarchy, must constantly reassert their dominance by physical means). Moreover, these links between rank and pathology can be made even more dramatic by the culture of a particular social group and by a personality prone toward interpreting ambiguous social circumstances as psychologically stressful.

Negative Physiological Effects of Stressful Social Ranking

Studies of both feral and captive animal populations show that animals with specific

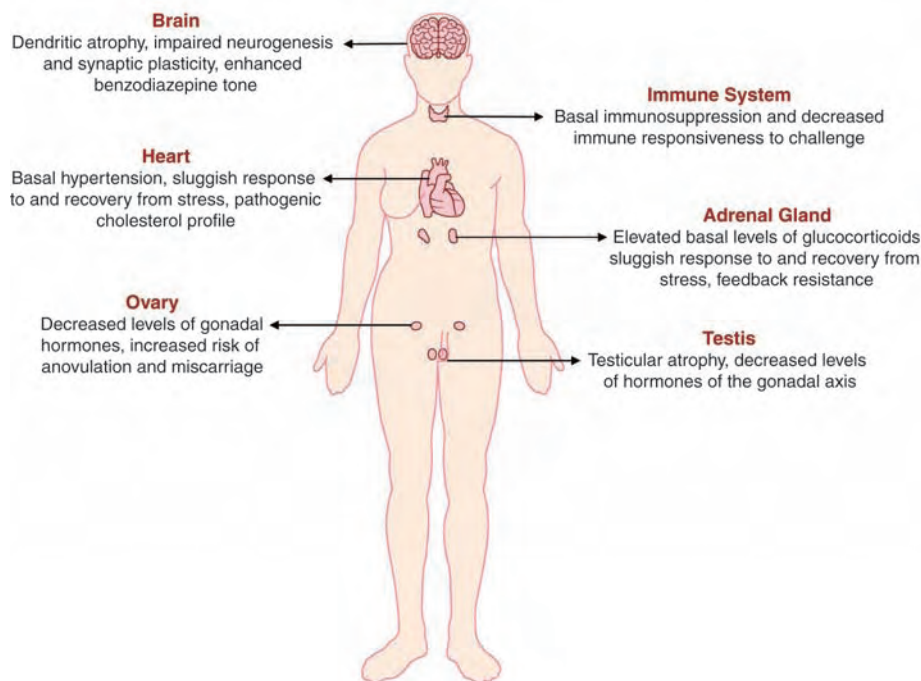


Fig. 2. Physiological correlates of the more stressful social rank. [Image credit: Bayard Colyear, Stanford Visual Arts Services]

dominance ranks tend to show characteristic stress-related physiological profiles (Table 1). We know that a particular rank gives rise to a particular physiological profile, rather than visa versa, because studies of individual captive animals before they are placed in social groups indicate that physiological profiles of singly-housed subjects do not predict their subsequent ranks in a social group (38).

Several stress-related physiological endpoints have been found to be sensitive to rank. The most frequently studied endpoint is the blood level of glucocorticoids (GCs), adrenal steroid hormones that are secreted during stress, such as cortisol or hydrocortisone in primates and corticosterone in many rodent species. GCs typify the double-edged nature of the stress response, as they help mediate adaptation to short-term physical stressors yet are pathogenic when secreted chronically.

Consistently, animals who are more socially stressed by the dominance hierarchy show indices of hyperactivity of the GC system. This includes elevated basal levels of GCs, the enlarged adrenal glands that accompany such increased secretion, a sluggish GC stress response in the face of a major homeostatic challenge, and impaired sensitivity of the system to negative feedback regulation.

In some cases, it is dominant individuals who show this profile. This includes species where dominant individuals have to repeatedly and physically reassert their rank (e.g., feral populations of dwarf mongooses, African wild dogs, female ring-tailed lemurs, and male chimpanzees) (12, 13, 39); those that are cooperative breeders (feral wolves and captive marmosets and tamarins) (16, 21); and those with transient periods of major rank instability (feral baboons and captive populations of talapoin, squirrel, and rhesus monkeys) (22).

In contrast, this profile is seen among subordinate individuals in species where high rank is maintained through nonphysical intimidation and the hierarchy is stable (feral male baboons and captive populations of squirrel and rhesus monkeys, tree shrews, rats, and mice) (22, 40, 41); where subordinates are exposed to frequent social stressors amid low availability of social support and minimal presence of kin (feral ring-tailed lemurs and captive populations of male rhesus or female talapoin monkeys) (13, 24); and when animals are in an enclosure too small to allow subordinate individuals to evade dominant ones (26).

A second prominent feature of the stress response is secretion of the catecholamine hormones (epinephrine and norepinephrine). These hormones of the sympathetic nervous system are secreted within seconds of the onset of a stressor (versus minutes for GCs) and have many of the same effects as GCs upon metabolism and cardiovascular tone. Thus, as with GCs, although the acute secretion of catecholamines is adaptive, prolonged secretion can be pathogenic. The

speed with which catecholamines are secreted typically precludes measuring basal circulating levels (because of the stress caused by the restraint of subjects for taking blood samples), and the hormones are poorly and variably preserved in urine and feces. Thus, little is known about rank-catecholamine relationships.

Prolonged stress adversely affects cardiovascular function, producing (i) hypertension and elevated heart rate; (ii) platelet aggregation and increased circulating levels of lipids and cholesterol, collectively promoting atherosclerotic plaque formation in injured blood vessels; (iii) decreased levels of protective high-density lipoprotein (HDL) cholesterol and/or elevated levels of endangering low-density lipoprotein (LDL) cholesterol; and (iv) vasoconstriction of damaged coronary arteries. A small literature demonstrates that animals who are more socially stressed by the dominance hierarchy demonstrate (i) basal hypertension; (ii) a sluggish activation of the cardiovascular stress response after a challenge and delayed recovery when it abates; (iii) a pathogenic cholesterol profile; and (iv) increased vulnerability to the atherogenic effects of a high-fat diet. These are traits of subordinate individuals when the dominance hierarchy is stable (among captive fascicularis macaques of both genders and among feral male savanna baboons) but of dominant individuals of the same populations when the hierarchy is unstable (37, 42, 43).

Chronic stress inhibits reproduction in both genders, a classic example of stress suppressing a costly anabolic process until more auspicious times. In females, this suppression can take the

form of delayed puberty, decreased levels of estrogen and progesterone, increased incidence of anovulatory cycles, impaired implantation, greater risk of miscarriage, prolonged interbirth intervals, and accelerated reproductive senescence. Primate studies show that the stress of subordination in a stable hierarchy (of cynomolgus monkeys) is associated with decreased gonadal hormone levels (42); there are conflicting data as to whether dominance or subordination in stable hierarchies of feral baboons is associated with higher rates of miscarriage (44, 45).

Among males, prolonged and major stress can suppress fertility; at an extreme in teleost fish, this includes atrophy of testes and of hypothalamic regions responsible for gonadotropin release (46). More commonly, stress can suppress circulating testosterone levels (9). However, there are many exceptions, as numerous species are resistant to this effect when the stressor is male-male competition during mating seasons; moreover, it is not clear how often these lower testosterone levels actually affect behavior or fertility. There is no consensus as to whether more socially stressed individuals have lower basal testosterone levels. However, such individuals (in this case, subordinate male baboons in a stable hierarchy) are more vulnerable to the suppressive effects of stress on basal testosterone levels (9).

Stress has complex time- and severity-dependent effects upon immunity. In general, mild to moderate transient stressors enhance immunity, particularly the first phase of the immune response, namely innate immunity. Later

Table 1. Influence of societal characteristics on stress experienced by high- and low-ranking individuals. An asterisk indicates no rank-related trend.

Societal characteristic	Individuals experiencing the most stress
<i>Dominance style and means of maintaining despotic dominance</i>	
Despotic hierarchy maintained through frequent physical reassertion of dominance	High-ranking
Despotic hierarchy maintained through intimidation	Low-ranking
Egalitarian hierarchy	*
<i>Style of breeding system</i>	
Cooperative	High-ranking
Competitive	*
<i>Stability of ranks</i>	
Unstable	High-ranking
Highly stable	Low-ranking
<i>Availability of coping outlets for subordinates</i>	
High availability	*
Low availability	Low-ranking
<i>Ease with which subordinates avoid dominant individuals</i>	
Easy avoidance	*
Difficult avoidance	Low-ranking
<i>Availability of alternative strategies to overt competition</i>	
Present	*
Lacking	Low-ranking
<i>Personality</i>	
Dominants perceive neutral interactions as challenging; subordinates take advantage of coping strategies	High-ranking
Dominants are adept at exerting social control and highly affiliative; subordinates are poor at exploiting opportunities for coping and support	Low-ranking

phases of the stress response are immunosuppressive, returning immune function to baseline. Should the later phase be prolonged by chronic stress, immunosuppression can be severe enough to compromise immune activation by infectious challenges (47, 48). In contrast, a failure of the later phase can increase the risk of the immune overactivity that constitutes autoimmunity. No studies have examined rank differences in the first immunostimulatory phase of the stress response or in the risk of autoimmunity if the later suppressive stage fails to occur. However, suppression of circulating lymphocyte numbers and blunted immune responsiveness to a challenge have been reported among animals socially stressed by a dominance hierarchy (subordinate rodents and pigs subject to high rates of attack and dominant chimpanzee males in an unstable captive population). Less clear is whether such rank effects are of sufficient magnitude to actually increase the risk of infectious disease (47, 49).

Animals who are socially stressed by the dominance hierarchy for prolonged periods undergo neurobiological changes as well. This can involve inhibition of neurogenesis, dendritic atrophy, and impairment of synaptic plasticity in the hippocampus (50, 51) and altered patterns of apoptotic cell death (increases in the cortex and decreases in the hippocampus) (52); these pathologies have been observed in socially subordinate rodents and tree shrews in stable hierarchies in captive populations.

Finally, a socially stressful position in a hierarchy is also associated with alterations in the neurochemistry of anxiety. Receptors exist in the nervous system for the anti-anxiety benzodiazepines (BDZs), which include the synthetic molecules diazepam and chlordiazepoxide hydrochloride as well as an as-yet uncharacterized endogenous BDZ. Pharmacological blockade of BDZ receptors caused the greatest disinhibition of anxiety-related behaviors in subordinate males in a stable hierarchy among feral baboons (34). This rank difference was interpreted as reflecting the demands for anxious vigilance among such individuals, necessitating a greater counteracting effect of endogenous BDZ tone.

Human Hierarchies and Health

The literature reviewed raises the obvious question: Are these findings relevant to humans? Initially, they seem to be of minimal relevance. Humans are not hierarchical in the linear, unidimensional manner of many species. For example, humans belong to multiple hierarchies and tend to value most the one in which they rank highest (for example, a low-prestige employee who most values his role as a deacon in his church). Furthermore, the existence of internal standards makes humans less subject to the psychological consequences of rank. Finally, health-rank relations that are easy to study can be highly artificial (e.g., ex-

amining the physiological consequences of winning versus losing an athletic competition).

Despite these caveats, the SES gradient of health among Westernized humans is a robust example of social inequalities predicting patterns of disease. As mentioned earlier, stepwise descent in SES predicts a major increase in the incidence of an array of diseases and mortality (1–4).

These health effects of SES are not a result of poverty causing limited access to health care. Robust SES-health gradients exist in countries with universal health care and documented equality of access. In addition, gradients exist for diseases with incidences that are impervious to preventative health measures (e.g., juvenile diabetes) (2, 3).

Only a small portion of the SES-health relationship is due to SES-related life-style differences. In Westernized societies, lower SES is associated with higher rates of smoking and drinking to excess, less healthy diets, more sedentary life-styles, crime- and toxin-riddled communities, and fewer coping outlets (e.g., health club memberships and vacations). However, the most prominent of these factors collectively account for only a small part of the variability in the SES-health gradient (3).

Instead, increasing evidence suggests that the gradient arises from psychosocial factors. Subjective SES can be at least as predictive of health as is objective SES (1); in other words, feeling poor may be at the core of why being poor predicts poor health. In the United States, at the level of states or cities, the same low SES predicts poorer health in communities with greater income inequality (4). Whereas large inequalities decrease the availability of protective life-style factors for the poor in a community (what has been termed a “neomaterialist” explanation for the inequality-health relationship) (53), the disease consequences of feeling poor are often rooted in the psychosocial consequences of being made to feel poor by one’s surroundings (4). Increased income inequality typically decreases a community’s “social capital” (shown in decreased levels of trust and increased senses of alienation and disenfranchisement), and such decreased capital mediates the relationship between income inequality and health (2).

Conclusions

Strong associations between social status and health thus occur in numerous species, including humans, with the poor health of those in the “wrong” rank related to their surfeit of physical and psychosocial stressors. In considering these issues in nonhuman species, the variability, qualifiers, and nuances of the rank-health relationship are frequently emphasized, a testament to the social complexity of other species. In contrast, in humans, there is a robust imperviousness of SES-health associations to differences in social and economic systems. It is not plausible that this human/nonhuman contrast re-

flects human sociality being less complex than in, say, baboons. Instead, it is a testimony to the power of humans, after inventing material technology and the unequal distribution of its spoils, to corrosively subordinate its have-nots.

References and Notes

1. N. Adler et al., *Health Psychol.* **19**, 586 (2000).
2. I. Kawachi, B. Kennedy, *The Health of Nations: Why Inequality Is Harmful to Your Health* (New Press, New York, 2002).
3. J. Siegrist, M. Marmot, *Soc. Sci. Med.* **58**, 1463 (2004).
4. R. Wilkinson, *Mind the Gap: Hierarchies, Health, and Human Evolution* (Weidenfeld and Nicolson, London, 2000).
5. E. Fehr, B. Rockenbach, *Curr. Opin. Neurobiol.* **14**, 784 (2004).
6. S. Brosnan, F. de Waal, *Nature* **425**, 297 (2003).
7. R. Sapolsky et al., *Endocr. Rev.* **21**, 55 (2000).
8. J. Brady et al., *J. Exp. Anal. Behav.* **1**, 69 (1958).
9. R. Sapolsky, *Psychoneuroendocrinology* **16**, 281 (1991).
10. B. Thierry, M. Singh, W. Kaumanns, Eds., *Macaque Societies: A Model for the Study of Social Organization* (Oxford Univ. Press, New York, 2004).
11. F. de Waal, *Chimpanzee Politics: Power and Sex Among Apes* (Johns Hopkins Univ. Press, Baltimore, MD, 1983).
12. S. Creel et al., *Nature* **379**, 212 (1996).
13. S. Cavigelli, *Anim. Behav.* **57**, 935 (1999).
14. R. Sapolsky, *Biol. Psychiatry* **28**, 862 (1990).
15. F. Bercovitch, A. Clarke, *Physiol. Behav.* **58**, 215 (1995).
16. D. Abbott et al., *Comp. Biochem. Physiol. C* **119**, 261 (1998).
17. D. Davis, J. Christian, *Proc. Soc. Exp. Biol. Med.* **94**, 728 (1957).
18. K. Manogue et al., *Primates* **16**, 457 (1975).
19. D. Abbott et al., *Comp. Biochem. Physiol. C* **119**, 261 (1998).
20. S. Schoech et al., *Physiol. Zool.* **70**, 68 (1997).
21. J. Sands, S. Creel, *Anim. Behav.* **67**, 387 (2004).
22. R. Sapolsky, in *Primate Social Conflict*, W. Mason, S. Mendoza, Eds. (SUNY Press, New York, 1993), pp. 171–183.
23. F. de Waal, *Science* **289**, 586 (2000).
24. D. Abbott et al., *Horm. Behav.* **43**, 67 (2003).
25. S. Creel, *Trends Ecol. Evol.* **16**, 491 (2001).
26. P. McLeod et al., *Can. J. Zool.* **74**, 209 (1996).
27. B. Smuts, *Sex and Friendship in Baboons* (Harvard Univ. Press, Princeton, NJ, 1999).
28. C. Virgin, R. Sapolsky, *Am. J. Primatol.* **42**, 25 (1997).
29. A. Maggioncalda et al., *Am. J. Phys. Anthropol.* **118**, 25 (2002).
30. J. Wingfield, R. Sapolsky, *J. Neuroendocrinol.* **15**, 711 (2003).
31. W. Goymann et al., *Horm. Behav.* **43**, 474 (2003).
32. D. Gust et al., *Brain Behav. Immun.* **5**, 296 (1991).
33. B. Wallner, in *The Integrative Neurobiology of Affiliation*, C. Carter, B. Kirkpatrick, I. Lederhendler, Eds. (New York Academy of Science, New York, 1996), pp. 45–51.
34. R. Sapolsky, L. Share, *PLoS Biol.* **2**, E106 (2004).
35. C. Virgin, R. Sapolsky, *Am. J. Primatol.* **42**, 25 (1997).
36. J. Ray, R. Sapolsky, *Am. J. Primatol.* **28**, 231 (1992).
37. S. Manuck et al., *Psychosom. Med.* **57**, 275 (1995).
38. D. Morgan et al., *Am. J. Primatol.* **52**, 115 (2000).
39. N. Masataka, *Ethology* **85**, 147 (1990).
40. S. Barnett, *Nature* **175**, 126 (1955).
41. J. Eberhart, *Physiol. Behav.* **30**, 361 (1983).
42. C. Shively, T. Clarkson, *Arterioscler. Thromb.* **14**, 721 (1994).
43. R. Sapolsky, L. Share, *Am. J. Primatol.* **32**, 261 (1994).
44. C. Packer et al., *Nature* **373**, 60 (1995).
45. J. Altmann et al., *Nature* **377**, 688 (1995).
46. M. Fox, R. Andrews, *Behaviour* **46**, 129 (1973).
47. R. Ader et al., *Psychoneuroimmunology* (Academic Press, San Diego, CA, 2001), ed. 3.
48. S. Cohen et al., *Psychosom. Med.* **59**, 213 (1997).
49. F. Dhabhar, B. McEwen, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1059 (1999).
50. Y. Kozorovitskiy, E. Gould, *J. Neurosci.* **24**, 6755 (2004).
51. A. Magarinos et al., *J. Neurosci.* **16**, 3534 (1996).
52. P. Lucassen et al., *Eur. J. Neurosci.* **14**, 161 (2001).
53. J. Lynch et al., *Milbank Q.* **82**, 5 (2004).
54. Research by the author was supported by grants from the MacArthur, the Harry Frank Guggenheim, and the Templeton Foundations and by the Office of the President, Republic of Kenya.

Horsfield's Hawk-Cuckoo Nestlings Simulate Multiple Gapes for Begging

Keita D. Tanaka* and Keisuke Ueda

Nestlings of some brood parasitic birds evict hosts' eggs and young soon after hatching, thereby avoiding discrimination by their hosts while monopolizing parental care (1, 2). However, eviction carries a cost, because lone parasitic nestlings attract a reduced provisioning rate (2, 3) and need to beg with supernormal signals (2). For example, in the case of the common cuckoo *Cuculus canorus*, a nestling begs with extremely intense begging calls to compensate for the deficient visual stimulus associated with the display of its single gape, which is smaller in total area than the gapes of a whole brood of host chicks (2), although the cuckoo chick itself is much larger than a host chick and requires proportionately more food.

With chicks of the evicting Horsfield's hawk-cuckoo *C. fugax*, the size relationship between host and parasite is similar to that of the common cuckoo, so chicks will also need to beg with additional signals. Here we describe a distinct form of signaling used by a Horsfield's hawk-cuckoo nestling to obtain sufficient food. The chick displays a gape-colored skin patch on the underside of each wing to host parents as they deliver food to the nest (Fig. 1A) (4). The display of the wing patch might simulate extra gapes and so stimulate increased provisioning. We tested this hypothesis experimentally (5).

If a hawk-cuckoo chick displays the wing patch as a begging signal, then the display ought to reflect its nutritional need (2, 6). We analyzed whether the frequency of display was affected by the chicks' nutritional need ($n = 6$ nestlings). The length of the interval between successive feedings was used as an index of their hunger condition. We found that cuckoo chicks were less likely to display a wing patch during begging as the interval between chick feedings became shorter, when chicks were presumably less hungry [generalized linear model (GLM), $\chi^2_1 = 20.1$, $P < 0.0001$] (Fig. 1B) (5). We suggest that a hawk-cuckoo chick solicits food from its host parents using the patch.

We performed a dyeing experiment in the field to test whether the wing patch actually functions to stimulate host parents to supply food. We set three experimental conditions and compared feeding rates by host parents under each: no manipulation (control I); applying a transparent solution to the wing patch (control II); and dyeing the wing patch black to cancel the signal (dyeing). We expected that the host

parents would reduce their provisioning rates when we dyed the patches black. As expected, the provisioning rates decreased only

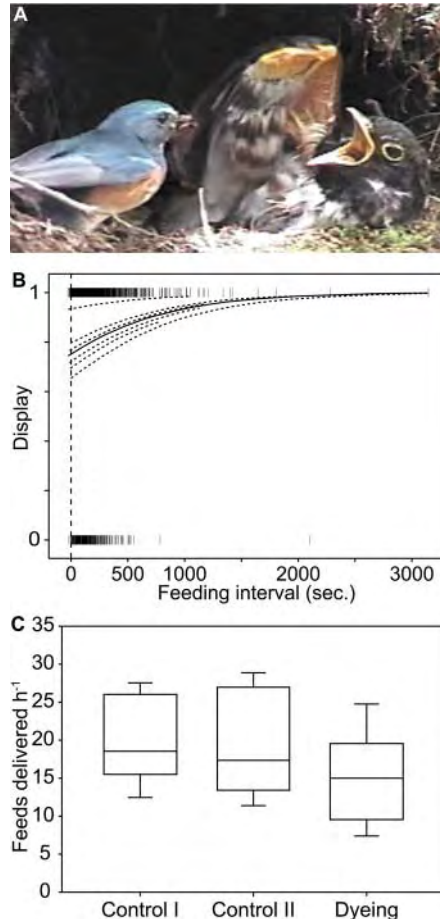


Fig. 1. (A) Display of the wing patch. A fledgling of the Horsfield's hawk-cuckoo raises and quivers the wing, exaggerating the wing patch as its host father delivers food. (B) Frequencies of wing patch display in relation to chicks' hunger. Each vertical bar shows the occurrence of the display according to the length of the related interval time: at 1, nestlings displayed the patch, and at 0, they didn't. The logistic curve (solid line) represents the pooled probability for these six nestlings that they will display the patch as the time they are kept unfed increases, and broken lines represent probabilities for individual nestlings (5). (C) Provisioning rates by host parents in relation to the experimental treatments for six nestlings (5). Boxes are composed of medians with first and third quartiles, and error bars show 10th and 90th percentiles of the distributions of hourly food deliveries in respective treatments.

under the dyeing condition (GLM, $\chi^2_1 = 6.93$, $P = 0.0085$) (Fig. 1C) (5). Human handling of nestlings had no significant effect on the outcome (control II, $\chi^2_1 = 1.21$, $P = 0.27$). We conclude that the Horsfield's hawk-cuckoo nestlings manipulate hosts to deliver sufficient food by displaying the wing patch.

Because host parents actually tried to place food mistakenly into the patch instead of a gape (4), a hawk-cuckoo chick is likely to induce overestimation of brood size by simulating a begging gape with the patch. Although a wing patch is not gape-shaped (Fig. 1A), it may be that host parents misperceive it as a gape because the inside of the nests built by three host species are typically dark (4). The decreased visibility in the dark nests may make host parents incapable of distinguishing between an actual gape and something that is gape-colored and moving like a begging chick.

Whereas the common cuckoo chick relies on vocal trickery, the Horsfield's hawk-cuckoo uses visual begging tricks to better exploit its hosts, perhaps because the nest sites of its hosts are more vulnerable to predators. Three host species of the hawk-cuckoo usually nest on the ground, and both parasite and host nestlings rarely beg loudly (7); this is likely an adaptation to intense predatory pressure by mammals because of the accessibility of the nests (8). In our study area, at least 50% of the nests were preyed upon each year. Host manipulation by the hawk-cuckoo nestlings may have evolved to stimulate host parents silently.

References and Notes

1. A. Lotem, *Nature* **362**, 743 (1993).
2. R. M. Kilner, D. G. Noble, N. B. Davies, *Nature* **397**, 667 (1999).
3. R. M. Kilner, J. R. Madden, M. E. Hauber, *Science* **305**, 877 (2004).
4. K. D. Tanaka, G. Morimoto, K. Ueda, *J. Avian Biol.*, in press.
5. Materials and methods are available as supporting material on Science Online.
6. R. M. Kilner, R. A. Johnstone, *Trends Ecol. Evol.* **12**, 11 (1997).
7. K. D. Tanaka *et al.*, unpublished data.
8. J. V. Briskie, P. R. Martin, T. E. Martin, *Proc. R. Soc. London Ser. B* **266**, 2153 (1999).
9. We thank G. Morimoto for assistance in the field and N. Yamaguchi and R. Kilner for comments. We conducted our fieldwork with the permission of the Forestry Agency and the Ministry of the Environment of Japan. Supported by Special Fund for Research of Rikkyo University and by Japan Society for the Promotion of Science grant no. 12640622.

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5722/653/DC1

Materials and Methods

Table S1

References and Notes

Movie S1

19 January 2005; accepted 14 March 2005

10.1126/science.1109957

Department of Life Sciences, Rikkyo University, 3-34-1 Nishi-Ikebukuro, Toshima, 171-8501 Tokyo, Japan.

*To whom correspondence should be addressed.
E-mail: keita@zaf.att.ne.jp

Structure of the Rotor of the V-Type Na⁺-ATPase from *Enterococcus hirae*

Takeshi Murata,¹ Ichiro Yamato,² Yoshimi Kakinuma,³
Andrew G. W. Leslie,^{4*} John E. Walker^{1*}

The membrane rotor ring from the vacuolar-type (V-type) sodium ion-pumping adenosine triphosphatase (Na⁺-ATPase) from *Enterococcus hirae* consists of 10 NtpK subunits, which are homologs of the 16-kilodalton and 8-kilodalton proteolipids found in other V-ATPases and in F₁F_o- or F-ATPases, respectively. Each NtpK subunit has four transmembrane α helices, with a sodium ion bound between helices 2 and 4 at a site buried deeply in the membrane that includes the essential residue glutamate-139. This site is probably connected to the membrane surface by two half-channels in subunit NtpI, against which the ring rotates. Symmetry mismatch between the rotor and catalytic domains appears to be an intrinsic feature of both V- and F-ATPases.

In eukaryotic cells, vital processes such as protein trafficking, endocytosis, neurotransmitter release, and intracellular pH regulation depend on the movement of ions across membranes by vacuolar adenosine triphosphatases (V-ATPases) (1), which are multisubunit complexes related to the F₁F_o- or F-ATPases found in eubacteria, mitochondria, and chloroplasts (2). Both classes have globular catalytic domains, V₁ and F₁, where ATP is hydrolyzed (and synthesized in F-ATPases), which are attached by central and peripheral stalks to the intrinsic membrane domains V_o and F_o, where ions are pumped across the membrane. ATP hydrolysis generates rotation of both the central stalk and an attached membrane ring of hydrophobic subunits. Ions are pumped through a pathway in the interface between the rotating ring and a static membrane component, which is linked to the outside of the V₁ or F₁ domain by the peripheral stalk.

Eukaryotic V-ATPases contain 13 different polypeptides (1). Subunits A to H form the V₁ domain. Subunits A and B are the counterparts of subunits β and α , respectively, in the F-ATPases; three copies of each subunit are arranged around the central stalk, which is made of single copies of subunits D and F.

The precise locations of the remaining subunits (C, E, G, and H) are uncertain, but they are thought to contribute to the peripheral stalk. The remaining subunits— a , c , c' , c'' , and d —form V_o. The closely related c , c' , and c'' subunits are each composed of tandem homologous repeats of the sequences of F-type c subunits (1). The F-type subunits are folded into two transmembrane α helices that are linked by a loop oriented to the same side of the membrane as the F₁ domain (3, 4). Their C-terminal α helices contain a conserved carboxylate side chain that is essential for ion transport (5). A carboxylate is also conserved in the equivalent position of helix 4 of the V-type subunits c and c' (in the second repeat) but not in helix 2 (the first repeat). Subunit c'' has a fifth nonessential transmembrane α helix at its N terminus, and its essential glutamate is in helix 3 (6, 7). In the V-ATPase from *Saccharomyces cerevisiae*, all three homologs are assembled into V_o and are essential for a functional enzyme (7); and, like the c subunits in F_o, they probably form the V_o rotary ring. V-ATPases also occur in prokaryotes; in *Enterococcus hirae*, a V-ATPase acts as a sodium ion extrusion pump (8). Its nine subunits are encoded in the Na⁺-dependent triphosphatase (*ntp*) operon. Subunits A, B, D, and G make up the V₁ domain and are the homologs of subunits β , α , γ , and δ , respectively, in mitochondrial F-ATPases. Subunit C of V-ATPase has no counterpart in F-ATPase and may connect the foot of the central stalk to V_o, which is made of subunits K and I, the homologs of the F-ATPase subunits c and a , respectively. The K subunit (NtpK) is also the homolog of the eukaryotic V-type c subunits (Fig. 1A). Like the c subunits in eukaryotic V-

ATPases, the sequence of NtpK contains tandem repeats in residues 1 to 79 and 80 to 156 with 22% identity and 44% similarity (Fig. 1B).

A surprising feature of F-ATPases is that the number of c protomers in the ring structures appears to vary between species. A complex of the F₁ domain with the c ring from the hydrogen ion-pumping H⁺-F-ATPase from *S. cerevisiae* contains 10 c protomers (4). Isolated c rings from spinach chloroplasts consist of 14 c protomers (9), and rings from the sodium ion-pumping Na⁺-F-ATPases from *Ilyobacter tartaricus* are made of 11 protomers (10). Biochemical and genetic evidence suggests that the H⁺-ATPases in *Escherichia coli* and *Bacillus* PS3 have 10 protomers per c ring (11, 12).

Here we describe the structure at 2.1 Å resolution of the K ring from *E. hirae*, which is the first high-resolution ring structure from either a V-type or an F-type enzyme.

Structure of the K ring. The structural data are summarized in Table 1. The 10 protomers of the K ring have very similar structures (Fig. 2); the root mean square deviation (RMSD) between α carbons is less than 0.06 and 0.08 Å for all atoms, except for crystal contact sites, where greater deviations occur. Each protomer is folded into five α helices, H0 to H4. H1 to H4 are transmembrane segments, and as expected from sequence homology, H1 and H2 resemble H3 and H4. The RMSD between all main-chain atoms in H1 and H2 (residues 11 to 46 and 51 to 79, respectively) compared with H3 and H4 (residues 87 to 122 and 127 to 155, respectively) is 1.0 Å. The 40 transmembrane α helices are packed in two concentric rings. H1 and H3 alternate in the inner ring, and H2 and H4 alternate in the outer ring. H1 and H3 (36 and 40 amino acids, respectively) are considerably longer than H2 and H4 (29 and 30 residues, respectively). H1 is linked to H2 and H3 to H4 by loop regions (residues 47 to 50 and 125 to 126, respectively), which, by analogy with F-ATPases, would lie proximal to the central-stalk region in the intact enzyme on the cytoplasmic side of the membrane. Another short loop (residues 80 to 84) links H2 and H3 on the periplasmic side of V_o in the intact enzyme. H0 (residues 1 to 8) also lies distal from V₁. It points inward and slightly upward at 61° to the axis of H1 toward the center of the ring. H2 is kinked around Pro⁶² and Gly⁶³, and H4 bends in the region of Met¹³⁷ and Val¹³⁸. The lower parts of H2 and H4 bend outward from the axis of the ring to approximately the same extent, as do the upper parts (mean values 11° and 26°, respectively); the bends in H1 and H3 are less pronounced. The effect of these curvatures is

¹The Medical Research Council Dunn Human Nutrition Unit, Hills Road, Cambridge CB2 2XY, UK.

²Department of Biological Science and Technology, Tokyo University of Science, Chiba, 278-8510, Japan.

³Faculty of Agriculture, Ehime University, Matsuyama 790-8566, Japan. ⁴The Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

*To whom correspondence should be addressed. E-mail: walker@mrc-dunn.cam.ac.uk (J.E.W.); andrew@mrc-lmb.cam.ac.uk (A.G.W.L.)

to widen the upper and lower parts of the ring relative to the central region, especially on the cytoplasmic side (external and internal diameters of 83 and 54 Å, respectively). On the opposite side of the membrane, the inward-pointing H0 regions narrow the internal diameter of the ring to 29 Å.

The height of the K ring (68 Å) is similar to that of the c rings from the F-ATPases from *S. cerevisiae* (58 Å) (4) and *I. tartaricus* (70 Å) (10). These similarities arise from the relationships between the sequences of the tandem repeats in subunit K with the sequence of the c subunit (Fig. 1B). For example, the first K repeat and the *S. cerevisiae* c subunit are 45% similar and 17% identical, respectively, and the second K repeat and c subunit are 52% similar and 25% identical, respectively. However, these similarities in sequence and structure of the V- and F-type protomers notwithstanding, their ring diameters differ greatly, and the V-type ring is much larger than the F-type ring (external and internal diameters at the top of the F-type ring in *S. cerevisiae* are 55 and 27 Å, respectively, and in *I. tartaricus*, the values are 50 and 17 Å; the corresponding values in the K ring are 83 and 54 Å) (fig. S1). These differences in diameter probably reflect significant differences in the association of the ring with the foot of the central stalk. In F-ATPases, the interaction between the ring and the central stalk subunits (γ , δ , and ϵ in the mitochondrial enzyme) is direct (4). In bacterial V-type enzymes, the C subunit is thought to be interposed between the equivalent central stalk subunits and the ring (13) (in eukaryotic V-ATPases, the homolog of bacterial subunit C is subunit d and not subunit C).

The Na⁺-binding sites. In each NtpK subunit, a strong peak of density between H2

and H4 adjacent to the γ carboxylate of Glu¹³⁹ was interpreted as a sodium ion. Other metal cations were also considered (Ca, Mg, Mn, Fe, Cu, Zn, and K), but Na⁺ was the only metal cation in the crystallization medium, and it fit the coordination geometry best (14). Sodium ions bind to the *E. hirae* V-ATPase with high affinity (15). Each sodium ion is surrounded by five oxygen atoms 2.2 to 2.3 Å distant, four of them in the side chains of Thr⁶⁴, Gln⁶⁵, Gln¹¹⁰, and Glu¹³⁹ (Fig. 3A), and the fifth in the main-chain carbonyl of Leu⁶¹. Thus, residues in H2, H3, and H4 all contribute to the Na⁺-binding pocket (Fig. 3). The bound sodium ions are occluded by the side chain of Glu¹³⁹, and the position of its γ carboxylate is stabilized by hydrogen bonds with the side chains of Gln¹¹⁰, Tyr⁶⁸, and Thr⁶⁴ (Fig. 3A). Access to the bound ion to

allow its exchange with the bulk phase (via a half-channel in subunit I, see below) can be achieved by changing the torsion angles of the side chain of Glu¹³⁹ without causing steric hindrance (Fig. 3C), although additional movements of adjacent side chains may also be necessary. Residues above and below the Na⁺-binding site are all hydrophobic, and there is no plausible intrinsic channel within the K ring itself to provide access to the binding site from either the cytoplasm or the periplasm.

Bound phospholipid, detergent, and water molecules. After the 10 NtpK protomers had been fitted into the electron density, many peaks of positive density around the ring remained uninterpreted. Twenty of them lying close to the inside surface are bifurcated, which suggests that they arise from the acyl side chains of phospholipids. They are dis-

Table 1. Data collection and crystallographic analysis. Values in parentheses indicate the highest resolution shell. The overall figure of merit (after density modification) is 0.247 (0.458).

Data set	Native 1	Native 2	K ₂ PtCl ₄	K ₂ PtCl ₆	KAuCl ₄
Wavelength (Å)	0.9792	0.9792	1.072	1.072	1.0397
Resolution range (Å)	61.3–2.10	67.4–2.30	87.7–2.80	67.4–3.20	67.4–2.90
R _{merge} (%) [*]	9.1 (54.5)	7.1 (52.5)	10.0 (69.8)	17.0 (70.6)	9.1 (42.8)
Completeness (%)	99.4 (97.5)	88.4 (64.0)	99.9 (99.8)	98.8 (93.3)	97.9 (86.8)
Unique observations	184,567	124,148	78,994	48,674	69,975
Redundancy	3.7 (2.9)	3.8 (3.1)	13.7 (8.8)	4.0 (4.0)	4.4 (3.8)
// σ	9.1 (1.8)	12.1 (2.3)	17.5 (2.9)	7.1 (2.1)	12.2 (2.7)
Wilson B factor (Å ²)	32.3	44.9	72.1	88.8	76.4
Phasing statistics					
Resolution range (Å)		20–2.30	20–3.0	20–3.5	20–3.0
Number of heavy atom sites	-	-	20	10	19
R _{cullis} [†] (centric)	-	-	0.653	0.907	0.899
Phasing power [‡] [iso/ano]	-	-	1.40/1.120	0.70	0.59

^{*}R_{merge} = $\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where h , k , and l are the Miller indices. [†]R_{cullis} = $\sum |F_{PH} - F_P| - F_H(\text{calc}) / \sum |F_{PH} - F_P|$, where $F_H(\text{calc})$ is the calculated heavy-atom structure factor. [‡]Phasing power = $F_H(\text{calc})/E$, where E is phase-integrated lack of closure.

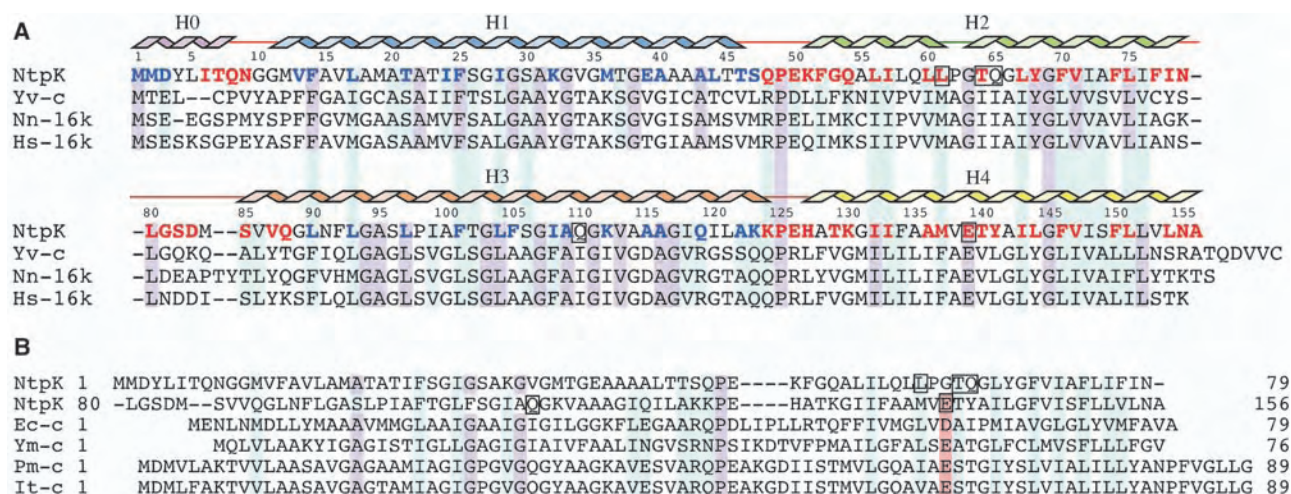


Fig. 1. Sequences of proteolipids from V- and F-ATPases (32). They were aligned with ClustalW (33). Dashes denote insertions. (A) The 16-kD proteolipids from V-ATPases from *E. hirae* (NtpK), the yeast vacuole (Yv-c), and human lysosomes (Hs-16k), and from gap junctions of *Nephrops norvegicus* (Nn-16k). The secondary structure, the positions of helices H0 to H4, and residue numbers of NtpK are shown above. Red and blue residues are exposed inside and outside the K ring, respectively. Violet and

light blue vertical boxes denote identical and similar residues, respectively. (B) The tandem repeats of NtpK aligned with c subunits of F-ATPases from *E. coli* (Ec-c), yeast mitochondria (Ym-c), *P. modestum* (Pm-c), and *I. tartaricus* (It-c). The N- and C-terminal residues are numbered on the left and right, respectively. Residues with conserved essential side chains are shaded red. Amino acids forming the Na⁺-binding site are boxed in (A) and (B).

tributed equally between the upper and lower regions of the model, with the openings of the bifurcation oriented downward in the upper layer and upward in the lower region, as expected in a phospholipid bilayer. The upper bifurcated moieties, but not the lower ones where the density is more limited, extend into regions of density that have been interpreted as phosphatidylglycerol (PG) head groups, because PG is the most abundant (57.1%) phospholipid head group in *E. hirae* (16). Because the most common fatty acyl chain in *E. hirae* phospholipids is C₁₆ (17), 10 1,2-dipalmitoyl-phosphatidylglycerol (DPPG) and 10 1,2-dipalmitoyl-glycerol (DPG) moieties were modeled into the upper and lower features, respectively (Fig. 4). The phosphate of the DPPG in the upper layer makes electrostatic interactions with Lys³² in H1 (O⁵-N^e distance 3.7 Å), and there is a hydrogen bond between O² in the glycerol moiety and Lys³² of an adjacent protomer (distance 3.4 Å) (Fig. 4B). One of the DPPG acyl chains interacts with H1 via the side chains of Leu¹⁷, Thr²¹, and Phe²⁵, and with H3 via the side chain of Phe¹⁰¹ in the adjacent protomer. The other DPPG acyl chain interacts with H1 via Phe²⁵ and with H3 via the side chains of Phe¹⁰¹ and Phe¹⁰⁵ (listed side chains have a buried surface area >15 Å²). The glycerol O⁹ of DPG in the lower layer forms a hydrogen bond with the main chain amido of Met² (distance 3.1 Å), and it also interacts with the α-amino group (distance 3.5 Å).

One of its acyl chains binds to H0 via the side chain of Met¹ and binds to Phe¹⁴ and Leu¹⁷ in H1 of the adjacent protomer (to the left) and to Leu⁹⁷ of the protomer two to the left. The other acyl chain binds to Met¹ and Met² in H0; Leu¹⁷ in H1; and Leu⁹⁰, Leu⁹³, and Leu⁹⁷ in the adjacent protomer to the left. The thickness of the internal hydrophobic region of the lipid bilayer (the separation between the lipid glycerol backbone region on the two sides of the membrane) is about 25 Å. In a lipid bilayer of di-(C16:0), this thickness is estimated to be about 26 Å (18).

Within the annulus of the bound phospholipid bilayer, additional regions of electron density are associated with the upper phospholipid leaflet but not the lower one. Thirty regions of elongated density were interpreted as acyl side chains (C₅ to C₁₀) in a second shell of phospholipids, although the possibility that they derive from detergent molecules cannot be excluded (Fig. 4, A and C). There was no significant electron density in the central part of the ring, which suggests that phospholipids in this region are mobile and disordered.

It is evident that all external bound phospholipids were lost during the isolation of K rings, because no bifurcated regions of density were associated with the outside of the ring. During crystallization, both dodecyl- and undecylmaltoside were present, and a number of features next to the external regions toward the top of H2 and H4 have been interpreted

as detergent molecules. The maltose ring is resolved in two molecules at crystal contact sites (Fig. 4A), and they appear to be undecyl- rather than dodecylmaltosides, but it is not possible to be certain. Twenty additional elongated features at approximately the same distance from the top of the ring have been interpreted as undecyl moieties, but the maltose moieties are disordered (Fig. 4, A to D). The model also contains 348 water molecules (Fig. 4) (16).

Coupling of the K ring and the central stalk. In the H⁺-V-ATPase from *Thermus thermophilus*, subunit C is thought to be at the foot of the central stalk (13). It is a pointed molecule made of 12 α helices surrounding a core of six additional α helices, all lying approximately parallel to a threefold axis of pseudosymmetry. It has been proposed that the tapered end of this bundle (diameter, 28 Å) penetrates 12 Å into the membrane ring (13). In *T. thermophilus*, the c protomers (subunit L) have two transmembrane α helices, similar to the F-type c subunits, and the ring was modeled as 12 protomers of the c monomer

Fig. 2. Structure of the K ring. The ribbon representations are shown in mono (A and B) and in stereo (C). H0 (residues 1 to 8), H1 (11 to 46), H2 (51 to 79), H3 (85 to 124), H4 (127 to 156), and loops (9 to 10, 47 to 50, 80 to 84, and 125 to 126) are colored violet, blue, green, orange, yellow, and red, respectively. The Na⁺ ions are shown as light blue spheres. (A) shows a view from the cytoplasmic side of the membrane; (B) and (C) show side views of the K ring and of the monomer, respectively. The cytoplasmic and periplasmic sides of the membrane are indicated.

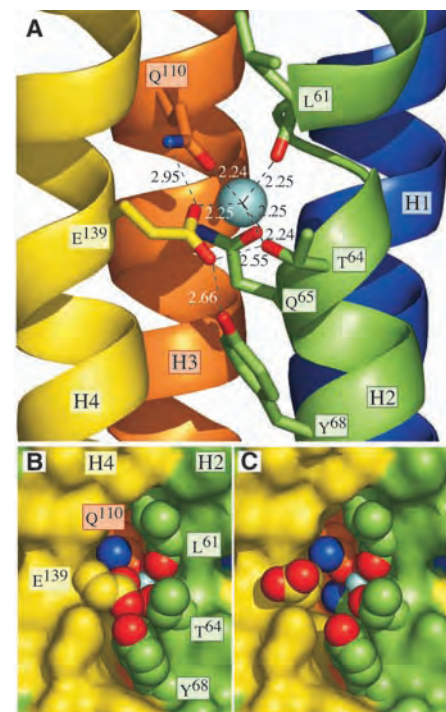
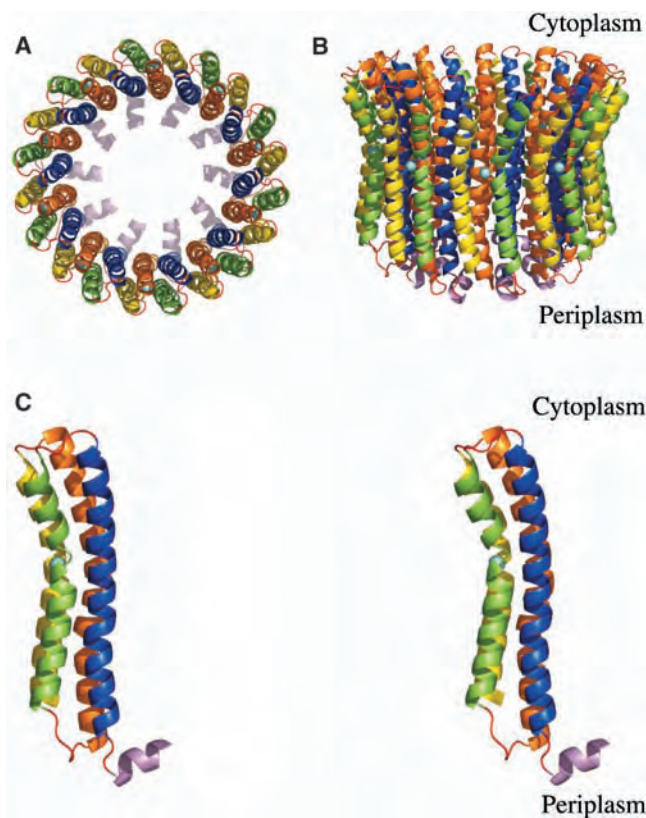


Fig. 3. The Na⁺-binding pocket viewed from outside the K ring. The color code for helices H1 to H4 and the Na⁺ ion is the same as in Fig. 2. Oxygen and nitrogen atoms are red and dark blue, respectively. Dotted lines are Na⁺-O bonds and potential hydrogen bonds, with distances in Å. Residues involved in Na⁺ binding are shown in stick (A) or space-filling (B and C) representation. In (B) and (C), the Na⁺ ion is in light blue, and residues not involved in binding it are represented as the solvent-accessible surface. In (C), to open the Na⁺-binding pocket, the torsion angles of the side chain of Glu¹³⁹ have been changed from (−170, 72), as determined in the structure, to (70, 170).

from *E. coli* F-ATPase determined by solution nuclear magnetic resonance (NMR) (3), giving a pore diameter of 30 Å, to act as a closely fitting socket for the 28 Å taper of subunit C. The sequence of NtpC is 20% identical and 40% similar to that of the C subunit of *T. thermophilus*, and it is likely to have a similar structure. The *T. thermophilus* C subunit structure was docked with the K ring, and, with the taper of subunit C penetrating into the central cavity almost to the narrowest region at the midpoint of the ring, it occupies most of the cavity, although some structural changes would be needed to achieve an extensive interaction (Fig. 4E). This model of the K ring with the C subunit differs substantially from the proposal (13) for the interaction of the C subunit with the c ring in *T. thermophilus*, where the smaller diameter of a ring assumed to be composed of 12 F-type c protomers restricts the penetration to the tip of the taper in subunit C.

Sodium ion translocation. Three key questions about the mechanism for transmembrane Na⁺ translocation are as follows: (i) How does the outer surface of the ring interact with the phospholipid bilayer. (ii) What is the nature of the interactions of the ring with subunit I? (iii) To what extent is the Na⁺-binding site accessible to the external milieu? Although no phospholipids are bound to the outer surface, their likely regions of interaction with the K ring can be inferred from features such as the distribution of charged and hydrophobic side chains and the locations of bound water molecules. In known structures of membrane proteins, tyrosine, asparagine, and

lysine residues (and also arginine, histidine, tryptophan, and serine) are often found near phospholipid head groups, acting as primary ligands for phosphodiester groups, and additional stabilization sometimes comes from threonine and glutamine residues (18, 19). The outside of the K ring has a hydrophobic belt 39 Å wide (Fig. 4D, α -carbon distance between Lys¹³⁰ and Asn¹⁵⁵) corresponding to the head-group separation in a lipid bilayer (18). This position for the lipid bilayer allows interactions between phosphodiester groups in the upper leaflet and the side chains of Gln⁵⁴, Thr¹²⁹, and Lys¹³⁰, and in the lower leaflet with Asn⁷⁹ and Asn¹⁵⁵. The distribution of bound water and detergent molecules supports this interpretation. This arrangement places the Na⁺-binding site close to the center of the bilayer, as in the c ring in *I. tartaricus* (10). Although several polar residues involved in binding the Na⁺ ion are exposed on the outer surface of the K ring, the network of hydrogen bonds surrounding the Na⁺ site satisfies all potential hydrogen-bonding groups (Fig. 3A). With a bound Na⁺, there is no energy penalty for these residues being near the center of the bilayer.

The position of the Na⁺-binding site close to the middle of the lipid bilayer, and the absence of any intrinsic K-ring channel leading to the site, support a “two-half-channel” ion translocation model (Fig. 5) as proposed for F-ATPases (20). The clockwise rotation of the K ring, as viewed from the cytoplasm, brings an occupied Na⁺-binding site into the K-ring-subunit-I interface. The proximity of the Na⁺ site to NtpI-Arg⁵⁷³, which is essential for ion

transport (21), produces an electrostatic interaction between Arg⁵⁷³ and Glu¹³⁹, disrupting the hydrogen-bond network and releasing the Na⁺ ion into the periplasm via a half-channel in NtpI (the ion release may possibly be assisted by electrostatic repulsion by R⁵⁷³). Further rotation, which may disrupt the Arg-Glu interaction, brings the site into contact with a second half-channel, resulting in the binding of a cytoplasmic Na⁺ ion. With the site occupied, the energy barrier to further rotation is lowered, bringing the site beyond the I-K-ring interface. This model differs from an earlier one for the V-ATPase based on Na⁺-exchange experiments in detergent (8, 15) where a single channel connected the ion-binding site to the periplasm at the I-K-ring interface, giving free cytoplasmic access to most Na⁺-binding sites. The earlier model is incompatible with the structure, and the kinetics of exchange (fast- and slow-exchange sites >1.7 and 0.16 min⁻¹, respectively) are much slower than physiological ion-pumping rates.

A “single channel” model of Na⁺ translocation has also been proposed for the F-ATPase from *Propionigenium modestum* (22), where the Na⁺-binding site is placed close to the membrane center on the basis of biochemical evidence (23, 24), but Na⁺-exchange experiments have been interpreted as suggesting free cytoplasmic access to the sites (25). The observations were reconciled by proposing an intrinsic c-ring channel connecting the ion-binding sites to the cytoplasm (10, 23, 2).

In a three-dimensional model of the c ring of the Na⁺-F-ATPase from *I. tartaricus* based

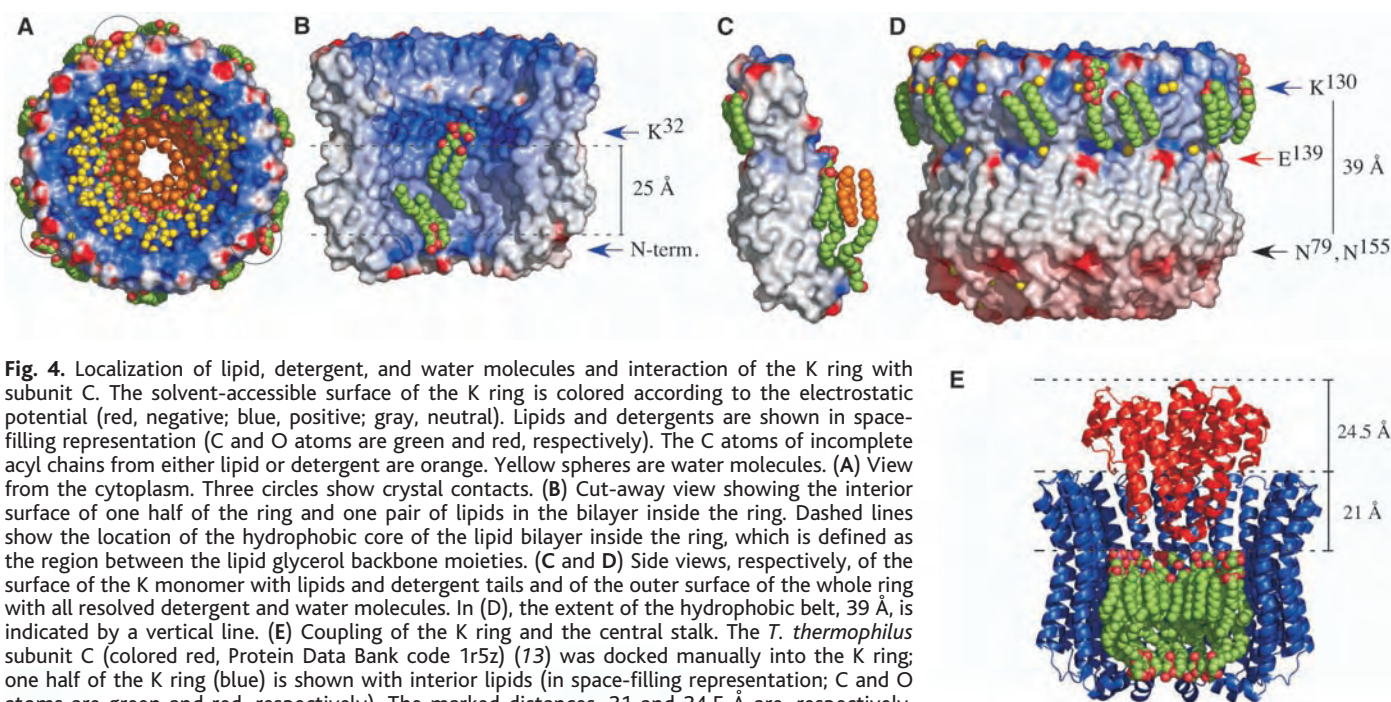


Fig. 4. Localization of lipid, detergent, and water molecules and interaction of the K ring with subunit C. The solvent-accessible surface of the K ring is colored according to the electrostatic potential (red, negative; blue, positive; gray, neutral). Lipids and detergents are shown in space-filling representation (C and O atoms are green and red, respectively). The C atoms of incomplete acyl chains from either lipid or detergent are orange. Yellow spheres are water molecules. (A) View from the cytoplasm. Three circles show crystal contacts. (B) Cut-away view showing the interior surface of one half of the ring and one pair of lipids in the bilayer inside the ring. Dashed lines show the location of the hydrophobic core of the lipid bilayer inside the ring, which is defined as the region between the lipid glycerol backbone moieties. (C and D) Side views, respectively, of the surface of the K monomer with lipids and detergent tails and of the outer surface of the whole ring with all resolved detergent and water molecules. In (D), the extent of the hydrophobic belt, 39 Å, is indicated by a vertical line. (E) Coupling of the K ring and the central stalk. The *T. thermophilus* subunit C (colored red, Protein Data Bank code 1r5z) (13) was docked manually into the K ring; one half of the K ring (blue) is shown with interior lipids (in space-filling representation; C and O atoms are green and red, respectively). The marked distances, 21 and 24.5 Å are, respectively, the depth of penetration of subunit C into the K ring and the extent to which subunit C protrudes from the ring and thus its contribution to the height of the central stalk.

on electron cryogenic microscopy (10), rods of density for the N- and C-terminal helices in an 11-fold ring and weaker connecting densities were observed, but side chains could not be identified. The model was constructed with a series of conserved glycines (residues 27, 29, 31, and 33) facing neighboring subunits in the inner ring of helices, with Gln³² (involved in Na⁺ binding) facing the C-terminal helix to achieve the tight packing of the inner helices suggested by conservation of the glycines. There was less information to constrain the orientation of the C-terminal helix, but because both Glu⁶⁵ and Ser⁶⁶ are implicated in Na⁺ binding (26), the face of the helix containing these residues was placed toward the N-terminal helix, allowing a chain of polar residues (Gln⁶¹, Thr⁵⁶, Ser⁵⁵, Asp⁵², and Lys⁵⁰) to connect the Na⁺-binding site to the cytoplasm, providing free access to the Na⁺-binding site.

To explore this proposal, a model of the *I. tartaricus* c ring based on the NtpK structure has been constructed (16). From this model, it appears that there is no intrinsic ion channel in the *I. tartaricus* ring. Thus, it is likely that Na⁺ translocation in the F-ATPases from *I. tartaricus* and *P. modestum* also relies on two half-channels (in subunit a) rather than on a single periplasmic channel, and this feature is common to ion transport mechanisms in both F- and V-ATPases.

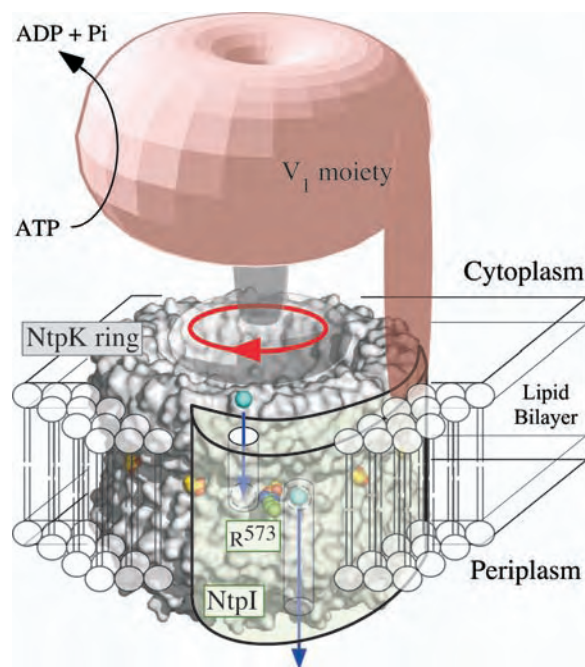
The overall fold of helices H3 and H4 of NtpK is very similar to the low-pH NMR structure of an *E. coli* c monomer (3). However, the position of the essential carboxylate (Glu¹³⁹ in NtpK and Asp⁶¹ in *E. coli* subunit c) differs by almost half an α -helical turn, corresponding to a rotation around the axis of the helix of approximately 150°. In view

of the sequence similarity and because Asp⁶¹ is assumed to be protonated in the NMR structure so that the ion binding site is occupied in both proteins, this result is surprising. A more detailed comparison of the NtpK and c-ring models will be described elsewhere.

Symmetry of rotor rings in V- and F-ATPases. By single-particle averaging of the same preparation of K rings as used for crystallography, a sevenfold symmetry was observed (27), which is incompatible with the structure of the K ring. A proper explanation of this discrepancy has not yet been found, but the x-ray structure demonstrates the 10-fold symmetry unequivocally, and it is likely that this is the symmetry of the ring in the fully assembled V-ATPase. Also, given the close sequence relationship between K subunits and c subunits in other V-ATPases, the 10-fold symmetry of the K rings is likely to have wider significance for V-ATPases. On the basis of various strands of evidence, including chemical analysis (28) and electron microscopy (29), it has become accepted that the symmetry of V-type rings is sixfold. This situation is reminiscent of the history of c-ring stoichiometries in F-ATPases, where, on the basis of similar kinds of experimentation, it was thought that the c rings had 12-fold symmetry, until higher resolution structural information demonstrated clearly the 10-fold symmetry in the F₁-c₁₀ complex from the F-ATPase from *S. cerevisiae* (4) and 11-fold and 14-fold symmetries in rings isolated from the Na⁺-ATPases from *I. tartaricus* (10) and the H⁺-ATPase from spinach chloroplasts (9), respectively. Thus, the earlier sixfold symmetries of c rings from V-ATPases will require reevaluation by high-resolution structural methods.

The fundamental importance of the observed ring stoichiometries is twofold. First, the range of values suggests that not all species pay the same energetic price either to make an ATP molecule in the F-ATPases or to pump a proton or a sodium ion in V-ATPases and bacterial F-ATPases. According to extant models for ATP synthesis, each 360° rotation of the central stalk and the attached c ring generates three ATP molecules in the F₁ domain, and the generation of a 360° rotation requires each of the c subunits in the ring to translocate an ion. Thus, the number of protons (or sodium ions) required to make an ATP molecule in enzymes with c ring symmetries of 10, 14, and 11 will be 3.3, 4.7, and 3.7, respectively. Similarly, hydrolysis of an ATP molecule in the V-ATPase from *E. hirae* will be coupled to the outward translocation of 3.3 sodium ions. Second, the mismatch of symmetries between catalytic and membrane domains observed now in both F- and V-ATPases could be an important intrinsic feature of both enzyme types, helping to reduce potential-energy barriers to rotation that might otherwise exist in enzymes with matching symmetries (4). Symmetry mismatch is also a feature of the machinery for packaging DNA into bacteriophage heads (30) and of the bacterial flagellar motor (31). Like the F- and V-ATPases, both machines have a rotary action.

Fig. 5. A model for ion translocation by the V-ATPase of *E. hirae*. The model is based on the structure of the K ring, shown in space-filling representation with Glu¹³⁹ in yellow with red oxygen atoms. Subunit I (structure unknown) is green and in close proximity to the K ring in space-filling representation with the essential residue Arg⁵⁷³ (27) in a half-channel connecting to the periplasm. Another half-channel links the ion-binding site to the cytoplasm (Na⁺ ions are blue). Clockwise rotation of the ring (arrow) driven by ATP hydrolysis in V₁ transfers bound Na⁺ ions, which have reached the ion-binding site via a half-channel connecting to the cytoplasm, to a position where they transfer to a second half-channel leading to the periplasm.



References and Notes

1. T. Nishi, M. Forgac, *Nat. Rev. Mol. Cell Biol.* **3**, 94 (2002).
2. J. E. Walker, *Angew. Chem. Int. Ed. Engl.* **37**, 2308 (1998).
3. M. E. Girvin, V. K. Rastogi, F. Abildgaard, J. L. Markley, R. H. Fillingame, *Biochemistry* **37**, 8817 (1998).
4. D. Stock, A. G. W. Leslie, J. E. Walker, *Science* **286**, 1700 (1999).
5. J. Hermolin, R. H. Fillingame, *J. Biol. Chem.* **264**, 3896 (1989).
6. T. Nishi, S. Kawasaki-Nishi, M. Forgac, *J. Biol. Chem.* **278**, 5821 (2003).
7. R. Hirata, L. A. Graham, A. Takatsuki, T. H. Stevens, Y. Anraku, *J. Biol. Chem.* **272**, 4795 (1997).
8. T. Murata, M. Kawano, K. Igarashi, I. Yamato, Y. Kakinuma, *Biochim. Biophys. Acta* **1505**, 75 (2001).
9. H. Seelert et al., *Nature* **405**, 418 (2000).
10. J. Vonck et al., *J. Mol. Biol.* **321**, 307 (2002).
11. W. Jiang, J. Hermolin, R. H. Fillingame, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4966 (2001).
12. N. Mitome, T. Suzuki, S. Hayashi, M. Yoshida, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12159 (2004).
13. M. Iwata et al., *Proc. Natl. Acad. Sci. U.S.A.* **101**, 59 (2004).
14. M. M. Harding, *Acta Crystallogr.* **D58**, 872 (2002).
15. T. Murata, K. Igarashi, Y. Kakinuma, I. Yamato, *J. Biol. Chem.* **275**, 13415 (2000).
16. Materials and methods are available as supporting material on Science Online.
17. F. A. Ibbott, A. Abrams, *Biochemistry* **19**, 2008 (1964).
18. A. G. Lee, *Biochim. Biophys. Acta* **1612**, 1 (2003).
19. H. Palsdottir, C. Hunte, *Biochim. Biophys. Acta* **1666**, 2 (2004).
20. W. Junge, H. Lill, S. Engelbrecht, *Trends Biochem. Sci.* **22**, 420 (1997).
21. M. Kawano, K. Igarashi, I. Yamato, Y. Kakinuma, *J. Biol. Chem.* **277**, 24405 (2002).
22. G. Kaim, U. Matthey, P. Dimroth, *EMBO J.* **17**, 688 (1998).
23. C. von Ballmoos et al., *J. Biol. Chem.* **277**, 3504 (2002).
24. C. von Ballmoos, T. Meier, P. Dimroth, *Eur. J. Biochem.* **269**, 5581 (2002).
25. G. Kaim, P. Dimroth, *EMBO J.* **17**, 5887 (1998).

26. G. Kaim, F. Wehrle, U. Gerike, P. Dimroth, *Biochemistry* **36**, 9185 (1997).
27. T. Murata *et al.*, *J. Biol. Chem.* **278**, 21162 (2003).
28. H. Arai *et al.*, *J. Biol. Chem.* **263**, 8796 (1999).
29. A. Holzenburg *et al.*, *Eur. J. Biochem.* **213**, 21 (1993).
30. R. W. Hendrix, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4779 (1978).
31. D. R. Thomas, D. G. Morgan, D. J. DeRosier, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10134 (1999).
32. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
33. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994).
34. This work was supported by the Medical Research Council. T.M. was supported partly by a Fellowship for Research Abroad from the Japan Society for the Promotion of Science. We thank the staff at the European Synchrotron Radiation Facility in Grenoble for their help. Coordinates and structure factors (accession codes 2bl2 and r2bl2sf) have been deposited in the Protein Data Bank.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1110064/DC1
Materials and Methods
SOM Text
Fig. S1
References

21 January 2005; accepted 8 March 2005
Published online 31 March 2005;
10.1126/science.1110064
Include this information when citing this paper.

Structure of the Rotor Ring of F-Type Na⁺-ATPase from *Ilyobacter tartaricus*

Thomas Meier,¹ Patrick Polzer,² Kay Diederichs,^{2*}
Wolfram Welte,² Peter Dimroth^{1*}

In the crystal structure of the membrane-embedded rotor ring of the sodium ion-translocating adenosine 5'-triphosphate (ATP) synthase of *Ilyobacter tartaricus* at 2.4 angstrom resolution, 11 c subunits are assembled into an hourglass-shaped cylinder with 11-fold symmetry. Sodium ions are bound in a locked conformation close to the outer surface of the cylinder near the middle of the membrane. The structure supports an ion-translocation mechanism in the intact ATP synthase in which the binding site converts from the locked conformation into one that opens toward subunit a as the rotor ring moves through the subunit a/c interface.

In the F₁F_o ATP synthase, the cytoplasmic F₁ catalytic domain (subunits $\alpha_3\beta_3\gamma\delta\epsilon$) is linked by means of a central and a peripheral stalk (subunits γ/ϵ and b_2/δ , respectively) to the intrinsic membrane domain called F_o (subunits ab_2c_{10-14}). Each of these domains functions as a reversible rotary motor and exchanges energy with the opposite motor by mechanical rotation of the central stalk. During ATP synthesis, energy stored in an electrochemical gradient of protons or Na⁺ ions fuels the F_o motor, which causes the stalk to rotate with the inherently asymmetric γ subunit acting as a camshaft to continuously change the conformation of each catalytic β subunit. These sequential interconversions, which result in ATP synthesis, endow the binding sites with different nucleotide affinities [for reviews see (1–3)]. The rotational model, which explains a wealth of biochemical and kinetic data, is impressively supported by the crystal structure of F₁ (4) and was experimentally verified by biochemical, spectroscopic, and microscopic techniques (5).

The F_o motor consists of an oligomeric ring of c subunits that is abutted laterally by the a and b₂ subunits (6). The c ring, together with γ/ϵ subunits, forms the rotor assembly, which spins against the stator components $ab_2\delta\alpha_3\beta_3$.

Ion translocation at the interface between subunit a and the c ring, driven by the ion motive force, is thought to generate torque (7–10) applied to the γ subunit, which is then used to promote the conformational changes required for ATP synthesis at the F₁ catalytic sites.

Despite intense efforts, little is known about the structural details of F_o. This lack of information hinders our understanding of how this molecular motor functions. The nuclear magnetic resonance (NMR) structures of the c monomer of *Escherichia coli* (7, 11) showed that the protein is folded into two α helices linked by a loop. Other structural studies indicated that the c subunits of the oligomer are tightly packed into two concentric rings of helices (12, 13). The number of c monomers per ring varies between $n = 10, 11$, and 14 units in the ATP synthases from yeast (12), *I. tartaricus* (13), and spinach chloroplasts (14), respectively.

Structure of the c ring. We chose to determine the crystal structure of the *I. tartaricus* c ring because of its inherent stability and relative ease of isolation (15). After purification and crystallization of wild-type c ring (16), the structure was solved (table S1) by molecular replacement (17) using a medium-resolution (6 Å) c-ring backbone model derived from electron crystallography (13). The asymmetric unit of the crystal contains 4 c rings. These rings are arranged in two parallel, but laterally translated, c-ring dimers each formed by a coaxial association of two rings that interact with their termini in a tail-to-tail fashion. A non-

crystallographic symmetry restraint was imposed during refinement at 2.4 Å over the 44 monomers in the asymmetric unit (3916 residues) with the exclusion of the loop regions that form crystal contacts, which substantially improved the electron density and resulted in an atomic model without Ramachandran plot outliers.

The electron density map of a single c ring shows a cylindrical, hourglass-shaped protein complex of ~70 Å in height, and with an outer diameter of ~40 Å in the middle and ~50 Å at the top and bottom. Eleven c subunits, each composed of two membrane-spanning α helices forming a hairpin, are arranged around an 11-fold axis, creating a tightly packed inner ring with their N-terminal helices (Fig. 1). The C-terminal helices pack into the grooves formed between N-terminal helices, producing an outer ring, in agreement with previous medium-resolution structures (12, 13). In the electron density map, the backbone and side chains of all amino acids are clearly defined, except for the C-terminal glycine. The N- and C-terminal helices are connected by a loop formed by the highly conserved peptide Arg⁴⁵, Gln⁴⁶, and Pro⁴⁷, which is exposed to the cytoplasmic surface (Fig. 2) (18, 19). The chain termini are exposed to the periplasm.

The C-terminal helices are shorter than the N-terminal helices, owing to a break at Tyr⁸⁰ followed by another short helix of one turn (Figs. 1 and 2). For each helix, an individual plane can be found that roughly contains the axis of the helix and the c-ring symmetry axis (Fig. 1A). All helices show a bend of about 20° in the middle of the membrane (at Pro²⁸ and Glu⁶⁵ in the N-terminal and the C-terminal helices, respectively), causing the narrow part of the hourglass shape. Moreover, the bend tilts the helices in the cytoplasmic half out of the plane by ~10°, yielding a right-handed twisted packing (Fig. 1A). When the c ring is viewed from the cytoplasm, it rotates counterclockwise during ATP synthesis (20) against the drag imposed by the F₁ motor components. Thus, the resulting torque might decrease the bend and increase the interhelical distance in the cytoplasmic part of the c ring, depending on the energies involved. Such a conformational change under load might serve to store elastic energy in the c ring, adding to that described for the central and peripheral stalk subunits (21). A change in the twist of the helices is supported by calculations (22) that show in the lowest-order mode a torsional

¹Institut für Mikrobiologie, Eidgenössische Technische Hochschule (ETH), Zürich Hönggerberg, Wolfgang-Pauli-Str. 10, CH-8093 Zürich, Switzerland. ²Fachbereich Biologie, Universität Konstanz M656, D-78457 Konstanz, Germany.

*To whom correspondence should be addressed. E-mail: dimroth@micro.bio.ethz.ch (P.D.); kay.diederichs@uni-konstanz.de (K.D.)

movement of the cytoplasmic against the periplasmic surface of the c ring (fig. S1).

Vonck *et al.* (13) determined the position of the periplasmic and the cytoplasmic interfaces of the membrane surrounding the c ring, which are approximately at Tyr⁸⁰ and Ser⁵⁵, respectively (Fig. 2). On the inner surface of the c ring, near the N- and C-termini, the map shows an extended electron density that can be modeled by an alkyl chain of nine C atoms per c subunit. We propose that detergents are bound here at a position corresponding to that of the external leaflet of the membrane. The Phe⁵ residues at the periplasmic end of the inner surface form a ring that possibly marks the position of the glycerol backbone of phospholipids. Such a positioning of lipids would correlate to the “central plug” feature seen in two-dimensional crystals of c rings (23). Despite the lack of direct evidence for a bilayer inside the c ring, its

existence can be inferred from the hydrophobic surface that extends beyond the electron density of the alkyl chains until Tyr³⁴ (fig. S2B). However, the internal bilayer appears to be wider and shifted toward the periplasmic surface with respect to its external counterpart.

Structure of the sodium ion binding site. The c ring was crystallized in buffer containing 100 mM sodium acetate, which promotes Na⁺ binding as shown by established techniques (15). The Na⁺ ions are seen in the map as 11 distinct densities at the bend of the helices, close to the outer surface of the c ring. Figure 2 shows a section of the c ring with two C-terminal and one N-terminal helix forming a Na⁺-binding unit. The coordination sphere is formed by side-chain oxygens of Gln³² (Oε1) and Glu⁶⁵ (Oε2) of one subunit and the hydroxyl oxygen of Ser⁶⁶ and the backbone carbonyl oxygen of Val⁶³ of the neighboring subunit

(Figs. 2 and 3). The distance between the liganding atoms and the ion is 2.37 ± 0.14 Å. The liganding residues are in good accordance with mutational studies, which recognized Gln³², Glu⁶⁵, and Ser⁶⁶ as being essential for Na⁺ binding (24), and their arrangement was confirmed as a Na⁺ binding site using the algorithm of Nayal and Di Cera (25). The crystal structure shows the bound ion surrounded by a network of hydrogen bonds: The Oε1 of Glu⁶⁵ accepts hydrogen bonds from the hydroxyl groups of Ser⁶⁶ and Tyr⁷⁰, and the Ne2 of Gln³² donates a hydrogen bond to Oε2 of Glu⁶⁵ (Fig. 3). These hydrogen bonds serve to keep Glu⁶⁵ deprotonated at physiological pH (26) in order to allow Na⁺ binding and to lock it into its ion-binding conformation. This arrangement of residues and their hydrogen bonds obviously serves to optimize the solvation energy (27) of the Na⁺ ion and results in a locked conformation of the ion binding site.

We propose that the present structure with the side chain of Tyr⁷⁰ facing outward and stabilizing the side-chain conformation of Glu⁶⁵ represents the conformation outside of the subunit a/c interface. Toward the periplasmic surface from the binding site, the structure forms a cavity (fig. S2B) to which the side chain of Tyr⁷⁰ could relocate to allow unloading and loading of the binding site to and from subunit a. This relocation may be part of an unlocking mechanism.

The Na⁺ binding signature is conserved in all known Na⁺-translocating ATP synthases and appears in the c subunits of other an-

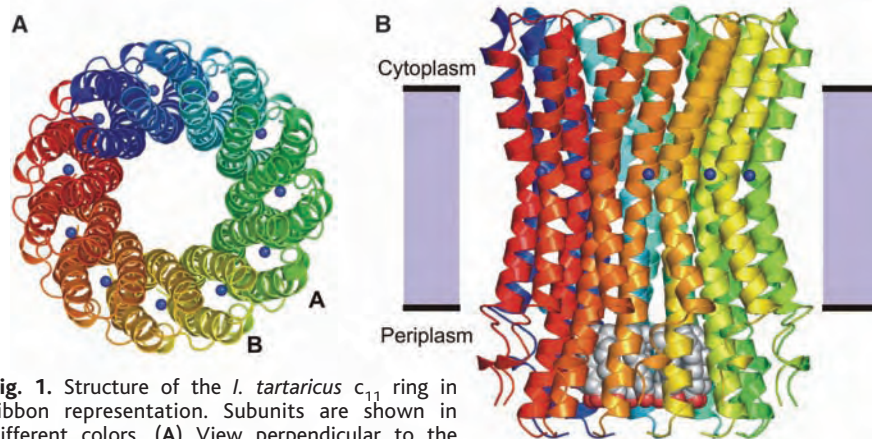


Fig. 1. Structure of the *I. tartaricus* c₁₁ ring in ribbon representation. Subunits are shown in different colors. (A) View perpendicular to the membrane from the cytoplasmic side. Two subunits are labeled. (B) Side view. The blue spheres represent the bound Na⁺ ions. Detergent molecules inside the ring are shown with red and gray spheres for clarity. The membrane is indicated as a gray shaded bar (width, 35 Å). The images were created with PyMOL (36).

Fig. 2. Section of the c ring showing the interface between the N-terminal and two C-terminal helices with those side chains discussed in the text. This three-helix bundle represents a functional unit responsible for Na⁺ binding and allowing access of the ion to the binding site. The view is normal to the external surface of the c ring with the ring axis approximately vertical. The color coding of the subunits is the same as in Fig. 1A.

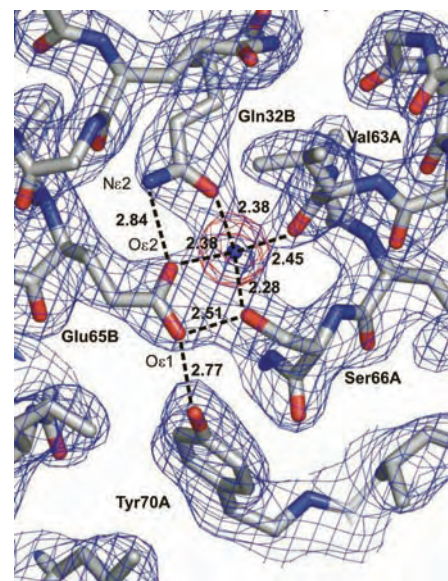
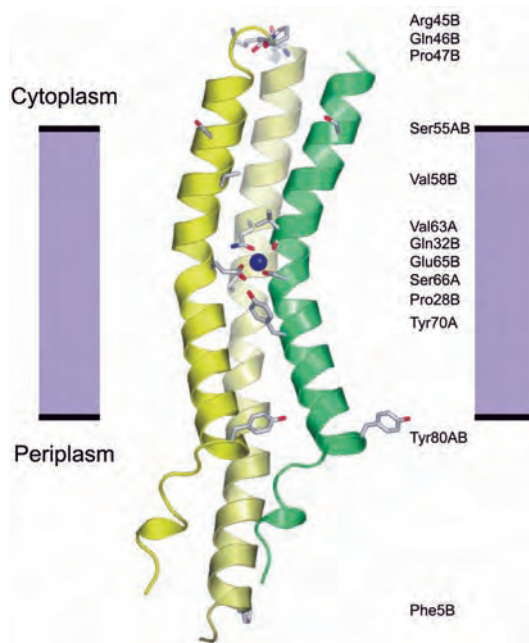


Fig. 3. Electron density map (red, Na⁺ omit map at 3.0 σ ; blue, $2F_{\text{obs}} - F_{\text{calc}}$ map at 1.4 σ) and residues of the Na⁺ binding site formed by two c subunits, A and B. Na⁺ coordination and selected hydrogen bonds are indicated with dashed lines. Distances are given in Å. The blue sphere indicates the center of the bound Na⁺ ion. The view is the same as in Fig. 2.

aerobic bacteria, suggesting that they may indeed harbor a Na⁺-ATP synthase (Fig. 4). Notably, none of these amino acids are conserved in proton-translocating ATP synthases, except for the acidic residue at position 65 (*I. tartaricus* numbering, Fig. 4), which is implicated in proton binding. Despite these differences in sequence, proton- and Na⁺-driven F₀ motors must share the same basic structure because of the demonstration that chimeric ATP synthases are functional (28).

Ion translocation in F₀ complexes. Previous models of the *E. coli* c ring were based on monomeric structures of the c subunit determined by NMR at pH 5 (11) and pH 8 (7) in solutions using organic solvents and were assumed to represent the protonated and unprotonated structures outside and inside the subunit a/c interface, respectively. In these models, the proton binding Asp⁶¹ (position 65 in *I. tartaricus*) of those c subunits, which are in contact with the lipids, faces toward the N-terminal helices. Within the subunit a/c interface, the C-terminal helix is proposed to swivel by ~140° clockwise (as viewed from the cytoplasm), which relocates Asp⁶¹ to become accessible to subunit a (29). The latter conformation is supported by cysteine cross-linking in F₀ complexes. These experiments indicate that residues at the same helical face as the proton binding site (Ile⁵⁵, Ala⁶², Met⁶⁵, Gly⁶⁹, Leu⁷², and Tyr⁷³) are accessible from subunit a and thus must be located on the surface of the c ring (30).

In the structure presented here, the placement of the Na⁺ binding site near the surface of the c ring between two C-terminal helices allows for ion transfer to and from subunit a after side-chain movements and does not require large rearrangements of the protein backbone. If the structure of the decameric *E. coli* c ring is modeled according to that of *I. tartaricus*, all residue positions that formed cross-links with subunit a are exposed to the surface (fig. S3). We therefore presume that the tertiary fold of the *E. coli* c monomer in the NMR experiments does not match that in the oligomeric c ring, and we note that the compact structure of the c ring would impose severe sterical restraints against the proposed swiveling.

In addition, the known specific reactivity of the ion-binding carboxyl group with the bulky dicyclohexylcarbodiimide (DCCD) could not be explained if the former is occluded between N- and C-terminal helices. With this specificity in mind, we inspected the c-ring surface near Glu⁶⁵, looking for a DCCD-accessible binding site. A pocket sufficiently large to accommodate binding of bulky organic molecules extends from Glu⁶⁵ to Val⁵⁸ at the ring surface within the interface between an N-terminal and two C-terminal helices (fig. S2A). The pocket is lined by small hydrophobic residues and is open to the membrane from where the hydrophobic DCCD molecule is thought to bind. The pocket might also contribute a binding site for other hydrophobic organic molecules such as

the new class of diarylquinoline antibiotics against tuberculosis that target subunit c of *Mycobacterium tuberculosis* (31).

To allow for unloading and loading of the ion binding site, the locked conformation must be converted to an open one during passage of the site through the subunit a/c interface. This interconversion may be enabled by modulating the electrostatic interactions of the binding site with the universally conserved Arg²²⁷ of subunit a (32) as the site passes this positively charged residue during rotation (fig. S4).

The path of the ion leading from the binding site into the cytoplasm is presently unknown. Two possibilities are discussed: a channel through subunit a (10) or through the c ring itself (3). In accord with the latter possibility, biochemical data suggest that the Na⁺ exit path is built intrinsically into the c ring (15). In the present c-ring structure, a channel leading from the binding site to the cytoplasmic surface could not be identified. Similarly, a channel-like pore is not visible in the structure of the Ca²⁺-ATPase (adenosine triphosphatase), which operates at a rate comparable to that of the ATP synthase (33), and in myoglobin (34). In contrast to these proteins with a low rate of ion flux, proteins with a structurally evident channel (e.g., potassium channel) have diffusion-limited transport rates up to 10⁸ ions per second (35), six orders of magnitude faster than the ATP synthase. However, dynamic fluctuations of the protein may open transient pathways for Na⁺ conduction so that a wider pore is not evident in the structure.

Taken together, the structural features of ion binding in the membrane-embedded c ring have profound implications for loading and unloading of the binding site because they represent stringent restraints for possible explanations of the F₀ motor function.

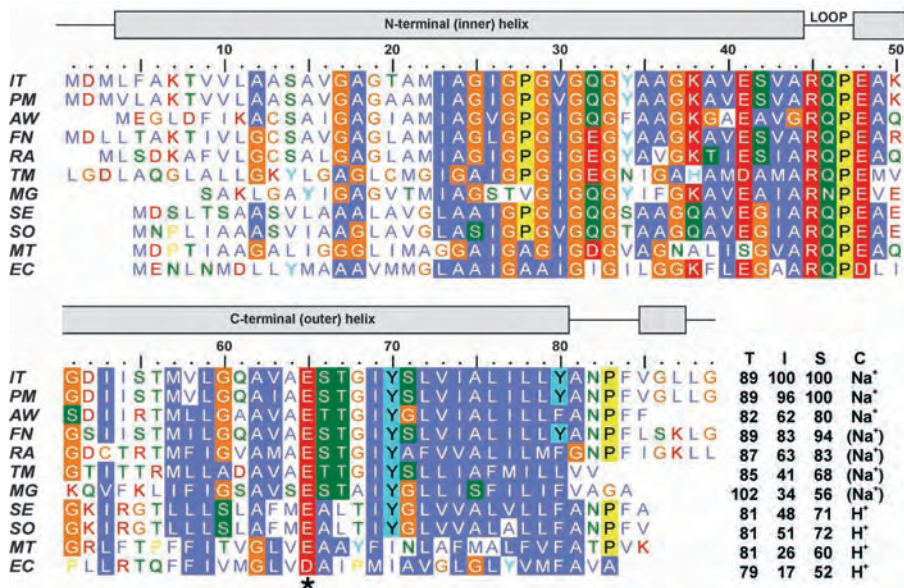


Fig. 4. Alignment of selected c subunit sequences. Amino acids are shaded with a threshold of 70% for identical and similar amino acids among the sequences shown. Orange: G, C; yellow: P; blue: A, V, L, I, M, F, W; green: S, T, N, Q; red: D, E, R, K; cyan: H and Y. Also shown are the total number of amino acids (T), the percentages of identical (I) and similar (S) amino acids (compared with *IT*), and the coupling ion (C) bound to the acidic residue marked by an asterisk. (Na⁺) marks putative Na⁺-binding c subunits. *I. tartaricus* (*IT*), *Propionigenium modestum* (*PM*), *Acetobacterium woodii* c2/3 (*AW*), *Fusobacterium nucleatum* (*FN*), *Ruminococcus albus* (*RA*), *Thermotoga maritima* (*TM*), *Mycoplasma genitalium* (*MG*), *Synechococcus elongatus* (*SE*), *Spinachia oleracea* (*SO*), *Mycobacterium tuberculosis* (*MT*), *Escherichia coli* (*EC*). Secondary-structure elements and numbering according to *IT* are indicated above the sequences.

References and Notes

- P. D. Boyer, *Annu. Rev. Biochem.* **66**, 717 (1997).
- R. A. Capaldi, R. Aggeler, *Trends Biochem. Sci.* **27**, 154 (2002).
- P. Dimroth, C. von Ballmoos, T. Meier, G. Kaim, *Structure (Camb.)* **11**, 1469 (2003).
- J. P. Abrahams, A. G. Leslie, R. Lutter, J. E. Walker, *Nature* **370**, 621 (1994).
- H. Noji, R. Yasuda, M. Yoshida, K. Kinosita Jr., *Nature* **386**, 299 (1997).
- J. L. Rubinstein, J. E. Walker, R. Henderson, *EMBO J.* **22**, 6182 (2003).
- V. K. Rastogi, M. E. Girvin, *Nature* **402**, 263 (1999).
- J. Xing, H. Wang, C. von Ballmoos, P. Dimroth, G. Oster, *Biophys. J.* **87**, 2148 (2004).
- A. Aksimentiev, I. A. Balabin, R. H. Fillingame, K. Schulten, *Biophys. J.* **86**, 1332 (2004).
- B. A. Feniouk et al., *Biophys. J.* **86**, 4094 (2004).
- M. E. Girvin, V. K. Rastogi, F. Abildgaard, J. L. Markley, R. H. Fillingame, *Biochemistry* **37**, 8817 (1998).
- D. Stock, A. G. Leslie, J. E. Walker, *Science* **286**, 1700 (1999).
- J. Vonck et al., *J. Mol. Biol.* **321**, 307 (2002).
- H. Seelert et al., *Nature* **405**, 418 (2000).
- T. Meier et al., *J. Mol. Biol.* **325**, 389 (2003).
- Materials and methods are available as supporting material on Science Online.
- L. C. Storoni, A. J. McCoy, R. J. Read, *Acta Crystallogr. D Biol. Crystallogr.* **60**, 432 (2004).
- S. D. Watts, Y. Zhang, R. H. Fillingame, R. A. Capaldi, *FEBS Lett.* **368**, 235 (1995).

19. J. Hermolin, O. Y. Dmitriev, Y. Zhang, R. H. Fillingame, *J. Biol. Chem.* **274**, 17011 (1999).
20. S. P. Tsunoda, R. Aggeler, M. Yoshida, R. A. Capaldi, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 898 (2001).
21. W. Junge *et al.*, *FEBS Lett.* **504**, 152 (2001).
22. K. Suhre, Y. H. Sanejouand, *Nucleic Acids Res.* **32**, W610 (2004).
23. T. Meier, U. Matthey, F. Henzen, P. Dimroth, D. J. Müller, *FEBS Lett.* **505**, 353 (2001).
24. G. Kaim, F. Wehrle, U. Gerike, P. Dimroth, *Biochemistry* **36**, 9185 (1997).
25. M. Nayal, E. Di Cera, *J. Mol. Biol.* **256**, 228 (1996).
26. W. R. Forsyth, J. M. Antosiewicz, A. D. Robertson, *Proteins* **48**, 388 (2002).
27. A. Warshel, J. Aqvist, *Annu. Rev. Biophys. Biophys. Chem.* **20**, 267 (1991).
28. G. Kaim, P. Dimroth, *Eur. J. Biochem.* **218**, 937 (1993).
29. R. H. Fillingame, C. M. Angevine, O. Y. Dmitriev, *FEBS Lett.* **555**, 29 (2003).
30. W. Jiang, R. H. Fillingame, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6607 (1998).
31. K. Andries *et al.*, *Science* **307**, 223 (2005).
32. F. Wehrle, G. Kaim, P. Dimroth, *J. Mol. Biol.* **322**, 369 (2002).
33. C. Toyoshima, M. Nakasako, H. Nomura, H. Ogawa, *Nature* **405**, 647 (2000).
34. H. Frauenfelder, G. A. Petsko, D. Tsernoglou, *Nature* **280**, 558 (1979).
35. D. A. Doyle *et al.*, *Science* **280**, 69 (1998).
36. W. L. DeLano, www.pymol.org (2002).
37. We thank the staff of the Swiss Light Source synchrotron site for support and M. S. Weiss for crystallographic advice. G. Cook and M. Koppenol are acknowledged for reading the manuscript and H.-J.

Apell is acknowledged for discussions. This work was supported by the Deutsche Forschungsgemeinschaft (TR SFB 11) and ETH Research Commission. The atomic coordinates and structure factors have been deposited in the Protein Data Bank (accession code 1YCE).

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5722/659/DC1

Materials and Methods

Figs. S1 to S4

Table S1

References

17 February 2005; accepted 31 March 2005
10.1126/science.1111199

Parietal Lobe: From Action Organization to Intention Understanding

Leonardo Fogassi,^{1,2*} Pier Francesco Ferrari,² Benno Gesierich,² Stefano Rozzi,² Fabian Chersi,² Giacomo Rizzolatti²

Inferior parietal lobule (IPL) neurons were studied when monkeys performed motor acts embedded in different actions and when they observed similar acts done by an experimenter. Most motor IPL neurons coding a specific act (e.g., grasping) showed markedly different activations when this act was part of different actions (e.g., for eating or for placing). Many motor IPL neurons also discharged during the observation of acts done by others. Most responded differentially when the same observed act was embedded in a specific action. These neurons fired during the observation of an act, before the beginning of the subsequent acts specifying the action. Thus, these neurons not only code the observed motor act but also allow the observer to understand the agent's intentions.

The posterior part of the parietal lobe has been traditionally considered as a typical association cortex. In this view, in the posterior parietal cortex, afferents from two or more sensory channels integrate, and this multimodal sensory association is the basis for some types of percepts, such as space. However, the works of Mountcastle (1) and Hyvärinen (2), and subsequent studies carried out in several laboratories (3–7), showed that the posterior parietal cortex, besides “putting together” different sensory modalities (association function), also codes motor actions and provides the representations of these motor actions with specific sensory information. This view stresses the importance of sensorimotor integration in the emergence of perception.

Recently, we reexamined the functional properties of the convexity of IPL (PF/PFG complex) in monkeys, testing the activity of single neurons in response to sensory stimuli and during monkey's active movements (8).

¹Dipartimento di Psicologia, Università di Parma, Borgo Carissimi 10, 43100 Parma, Italy. ²Dipartimento di Neuroscienze, Università di Parma, via Volturno 39, 43100 Parma, Italy.

*To whom correspondence should be addressed.
E-mail: fogassi@unipr.it

As previously reported (2), we found that IPL convexity has a motor somatotopic organization. Most interestingly, we found that many IPL neurons discharge both when the monkey performs a given motor act and when it observes a similar motor act done by another individual; these neurons are known as parietal mirror neurons (9–11). Here, we present data on the motor organization of IPL and on its mirror properties.

The coding of motor acts in the inferior parietal lobule. Neurons were recorded from the rostral sector of IPL (Fig. 1A) in two monkeys. All studied neurons ($n = 165$) were active in association with grasping movements of the hand (grasping neurons). They were formally tested in two main conditions. In the first condition, the monkey, starting from a fixed position (Fig. 1B, left), reached for and grasped a piece of food located in front of it and brought the food to the mouth (Fig. 1B, right, I). In the second condition, the monkey reached for and grasped an object, located as described above, and placed it into a container (Fig. 1B, right, II). In the first condition, the monkey ate the food brought to the mouth; in the second it was rewarded after correct accomplishment of the task.

Although some neurons discharged with the same strength regardless of the motor act that followed grasping, the large majority were influenced by the subsequent motor act. Examples are shown in Fig. 1C. Unit 67 discharged during grasping when grasping was followed by bringing the food to the mouth. In contrast, its discharge was virtually absent when grasping was followed by placing. Unit 161 exemplifies the opposite behavior. This neuron discharged very strongly when grasping was followed by placing, whereas only a weak discharge was present when grasping was followed by bringing to the mouth. Finally, Unit 158 did not show any significant difference in discharge intensity in the two conditions.

Table 1A summarizes the behavior of all recorded neurons. About two-thirds of neurons discharged preferentially ($P < 0.05$) when grasping was embedded into a specific motor action (12). The neuron selectivity remained unmodified when the conditions were blocked, as in Fig. 1, or interleaved (fig. S1).

Figure 1D shows the intensity and time course of neuron discharge of the grasp-to-eat and grasp-to-place populations in the two basic conditions. The population analysis (12), based on all selective neurons recorded from monkey M2, confirmed the data observed in individual neurons.

To control for the possibility that the differential discharge of neurons during the same motor act could be due to differences in the stimuli that the monkeys grasped, monkeys were trained to grasp an identical piece of food in both conditions. Food placing was achieved by showing the monkey a piece of food that the monkey particularly liked before each trial. After correct placing, the monkey received the preferred food. Unit 122 (Fig. 2) demonstrates that neuron selectivity did not depend on the stimulus used. The same result was found in all tested neurons ($n = 28$), regardless of whether they coded grasping to eat ($n = 22$) or grasping to place ($n = 6$) (12).

It is well known from human studies that the first motor act of an action is influenced by the next acts of that action (13). We also found that reaching-to-grasp movement followed by arm flexion (bringing the food to

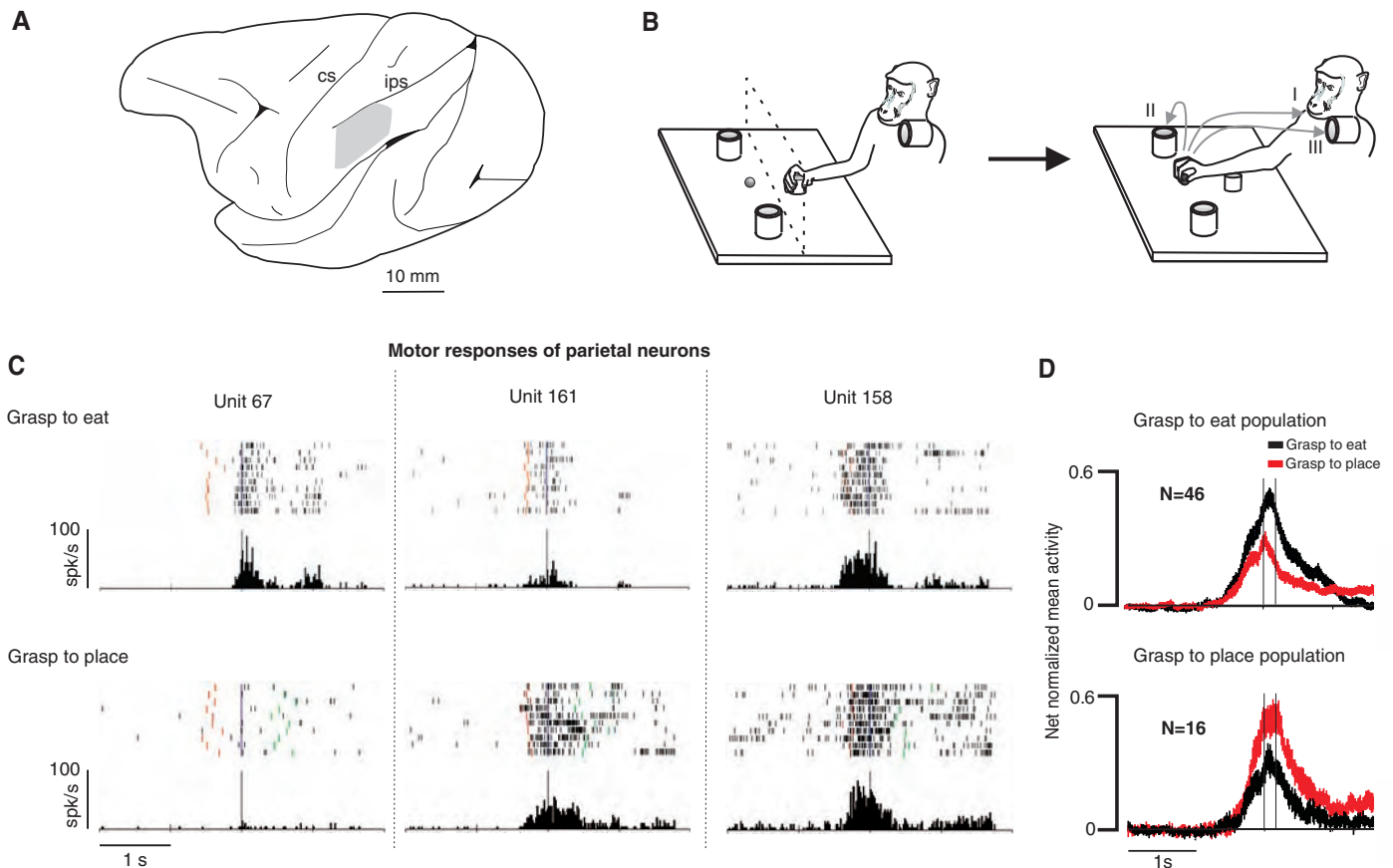


Fig. 1. (A) Lateral view of the monkey brain showing the sector of IPL (gray shading) from which the neurons were recorded. cs, central sulcus; ips, inferior parietal sulcus. (B) The apparatus and the paradigm used for the motor task. (C) Activity of three IPL neurons during grasping in conditions I and II. Rasters and histograms are synchronized with the moment when the monkey touched the object to be grasped. Red bars, monkey releases the hand from the starting position; green bars, monkey

touches the container; x axis, time, bin = 20 ms; y axis, discharge frequency. (D) Responses of the population of neurons selective for grasping to eat and grasping to place tested in conditions I and II. The two vertical lines in the two panels indicate the moment when the monkey touched the object and the moment in which the grasping was completed, respectively. The y axes are in normalized units. [For description of population analysis, see (12).]

the mouth) was executed significantly faster than the same movement followed by arm abduction (placing the food into the container), the wrist peak velocity being 68.4 and 62.8 cm/s, respectively ($t = 2.75$, $P < 0.05$). To render the reaching-to-grasp movement for placing kinematically more similar to that for eating, we introduced a third experimental condition. In this condition, the monkey was trained to grasp a piece of food (or an object) and to place it into a container located near the monkey's mouth (Fig. 1B, III). In terms of the goal, this new placing condition was identical to the previous one, but it required an arm flexion, as in the bringing-to-the-mouth condition. The kinematic analysis showed that the wrist peak velocity (74.8 cm/s) in the new placing condition was faster not only than that in the original placing condition ($t = 6.45$, $P < 0.0001$) but also than that in the grasping-to-eat condition ($t = 2.31$, $P < 0.05$). [For further kinematics data, see (12) and fig. S2].

The analysis of neurons studied in the three conditions ($n = 18$) showed that the grasping selectivity was independent of kinematics pa-

rameters. All neurons that discharged best during the grasping-to-eat condition ($n = 14$) discharged less in the grasping-to-place condition, both when placing was done in the container located near the target (basic condition, wrist velocity lower than in the grasping-to-eat condition) and when placing was done in the container located near the mouth (new condition, wrist velocity greater than in the grasping-to-eat condition). An example is shown in Fig. 2 (Unit 43). Similarly, all neurons that discharged best during grasping to place in the basic condition ($n = 4$) maintained the same selectivity when placing was done in the container located near the mouth, despite the different movement kinematics in the two conditions. The main factor that determined the discharge intensity was, therefore, the goal of the action and not the kinematics [see (12) for more details].

To control whether the force exerted by the monkey to grasp the objects placed at the target location could be responsible for the differential activation of IPL neurons, a mechanical device that could hold the objects with two different strengths was used. Twelve

neurons were tested with this device. No change in neuron selectivity was found in these neurons when the monkey grasped the objects using the two different forces (12).

Visual properties of mirror neurons in the inferior parietal lobule. In the IPL, there are neurons endowed with mirror properties. We studied 41 mirror neurons, all discharging both during grasping observation and grasping execution, in a visual task in which the experimenter performed, in front of the monkey, the same actions that the monkey did in the motor task, that is, grasping to eat and grasping to place (12).

Some neurons discharged with the same strength regardless of the motor act following the observed grasping. The majority of neurons, however, were differentially activated depending on whether the observed grasping was followed by bringing to the mouth or by placing. Examples are shown in Fig. 3. Unit 87 discharged vigorously when the monkey observed the experimenter grasping a piece of food, provided that he subsequently brought the food to his mouth. In contrast, the neuron's discharge was much weaker when, after grasping an object, the ex-

perimeter placed it into the container. Unit 39 illustrates the opposite behavior. Finally, Unit 80 did not show any significant difference in discharge intensity in the two conditions.

Table 1B summarizes the behavior of all tested mirror neurons. As in the motor conditions, the majority of neurons (75.6%) were influenced by the final goal of the action. Neurons responding to the observation of grasping to eat constituted the most represented neuron category.

The difference in discharge observed in the two basic conditions of the experiment

could be, in principle, attributed to the two different types of objects (food or solids) grasped by the experimenter. A series of neurons ($n = 20$) was thus tested in a modified version of the visual task consisting now of three conditions: grasping food to eat, grasping food to place, and grasping solid to place. For all neurons, the selectivity remained the same regardless of which object (food or solid) was grasped by the experimenter (12). Twelve of the tested neurons did not show any difference when the experimenter grasped food or solid to place (Fig. 4, Unit 126), whereas eight

neurons, all selective for the observation of grasping to eat, discharged more strongly in the placing conditions when the grasped object was food (Fig. 4, Unit 109).

As mentioned above, all 41 neurons tested with the visual task discharged during both grasping observation and grasping execution. In 19 of them, all motorically selective for one type of action (15 grasping-to-eat neurons and 4 grasping-to-place neurons), we studied their higher order visuomotor congruence by comparing their visual and motor selectivity during grasping-to-eat and grasping-to-place actions. The great majority of the tested neurons (16 out of 19) showed the same specificity during grasping observation and grasping execution (13 out of 15 and 3 out of 4 in the case of grasping to eat and grasping to place, respectively). Figure 5A shows an example of a congruent mirror neuron (Unit 169). The analysis of population-averaged responses of all neurons, comparing their visual and motor selectivity, is presented in Fig. 5B (12).

The coding of actions in the inferior parietal lobule. IPL neurons discharge in association with specific motor acts. The present study shows that most of them code the same act (grasping) in a different way according to the final goal of the action in which the act is embedded. Control experiments showed that this selectivity is not due to spurious factors such as the force exerted to grasp the object or differences in movement kinematics during object grasping determined by the subsequent arm movements.

Similarly, other factors, such as motivation and attention, that are known to influence the activity of the posterior parietal neurons (1, 2, 7) cannot explain the present findings. First, the type of reward the monkey received was the same in both grasping-to-eat and grasping-to-place conditions. Second, in the same experimental session we found intermixed neurons

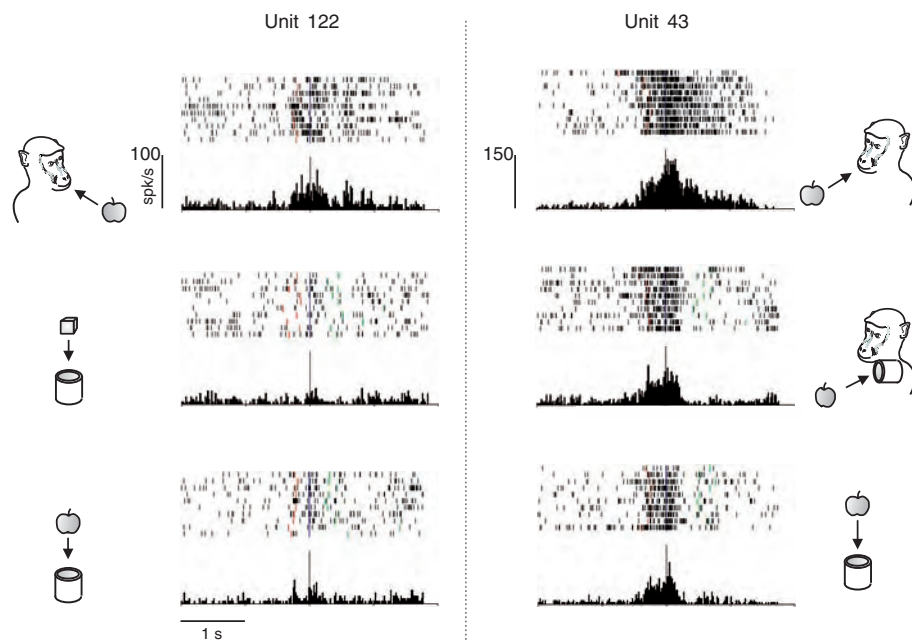


Fig. 2. Discharge of two IPL neurons during active grasping. Unit 122 strongly discharges when the monkey grasps a piece of food to eat (top), whereas it does not respond when the monkey grasps an object (center) or a piece of food (bottom) to place. Unit 43 strongly discharges when the monkey grasps a piece of food to eat (top), whereas the discharge is significantly weaker (12) when the monkey grasps a piece of food to place into a container positioned near the mouth (center) or near the grasped object location (bottom). Conventions as in Fig. 1.

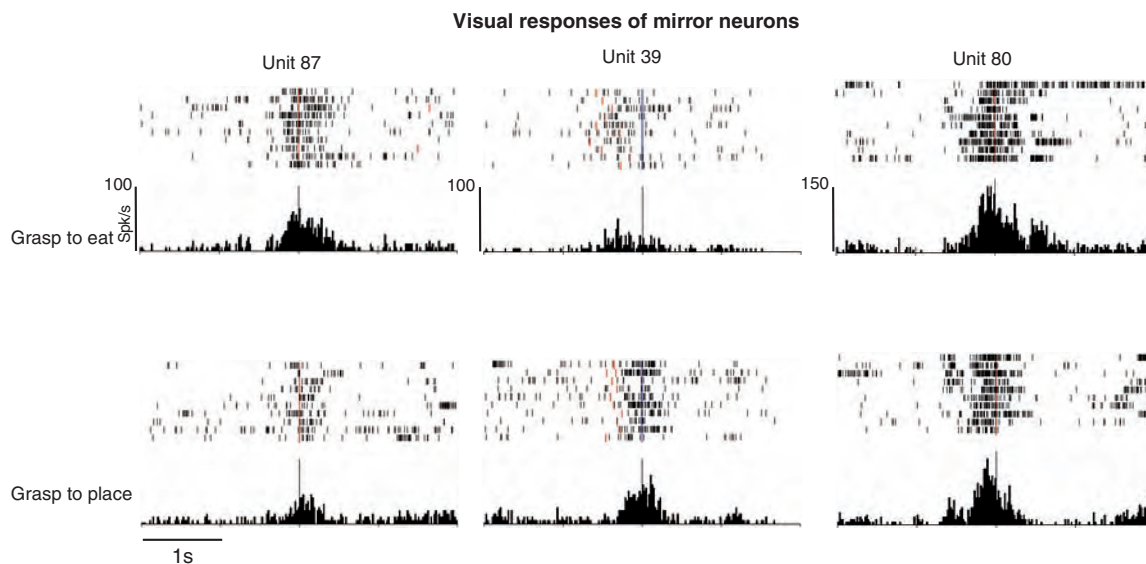


Fig. 3. Visual responses of IPL mirror neurons during the observation of grasping to eat and grasping to place done by an experimenter. Conventions as in Fig. 1.

showing specificity for grasping to place and for grasping to eat. Third, in a few experiments we alternated trials in which the monkey had to grasp to eat with others in which the monkey had to grasp to place. The neuron selectivity remained the same as during blocked trials (fig. S1).

The finding that most IPL neurons code the same motor act differently depending on the final action goal may appear counter-intuitive. From an engineering point of view, it should be more economical to have a multipurpose system of neurons for grasping, and

for other motor acts, that can be used whenever necessary. This would allow the possibility of having a smaller number of neurons coding the same motor act. There is, however, another aspect of motor organization that must be taken into consideration. One of the most striking aspects of action execution in animals is the fluidity with which the different motor acts follow one another. This “kinetic melody” (14) requires a close link between the different motor acts forming the entire action so that its execution can occur

without any gap. The system we found in the parietal lobe appears to have exactly this function. Motor acts are not related to one another independent of the global aim of the action but appear to form prewired intentional chains in which each motor act is facilitated by the previously executed one (15).

The organization of IPL receptive fields also favors a model that postulates a chain between neurons coding subsequent motor acts. Frequently, IPL neurons that respond to the passive flexion of the forearm have tactile receptive fields located on the mouth, and some of them respond in addition during grasping actions of the mouth (16). These neurons appear to facilitate the mouth opening when an object is touched or grasped by the monkey’s hand. Recently, several examples of this predictive chained organization in IPL have been described (7).

There are no available data on the possible specificity of ventral premotor cortex neurons in coding the same motor act according to the action to which it belongs. Two series of considerations suggest, however, that the same chained organization also exists in this region. First, the IPL is anatomically connected directly with the ventral premotor cortex (17–19). Thus, it is hard to believe that the parietal specificity would be lost in the next motor station. Second, and most important, the receptive field organization of the ventral premotor cortex shows characteristics similar to that just mentioned for the IPL (20).

Reading the intention of others: The inferior parietal mechanism. When an individual starts the first motor act of an action, he or she has clearly in mind what the final goal of the action is to which the motor act belongs. In the present study, the monkey’s decision on what to do with the object is undoubtedly made before the onset of the grasping movement. The action intention is set before the beginning of the movements and is already reflected in the first motor act.

This motor reflection of action intention and the chained motor organization of IPL neurons have profound consequences on a fundamental cognitive capacity, that of understanding the intention of others.

It is generally accepted that the fundamental role of mirror neurons is to allow the observing individual to understand the goal of the observed motor act (21–23). The rationale of this interpretation is the following: Because the monkey knows the outcome of the motor act it executes, it recognizes the goal of the motor act done by another individual when this act triggers the same set of neurons that are active during the execution of that act.

The finding of the present experiment is that IPL mirror neurons, in addition to recognizing the goal of the observed motor act, discriminate identical motor acts according to the action in which these acts are embedded. Because the discriminated motor act is part of a chain lead-

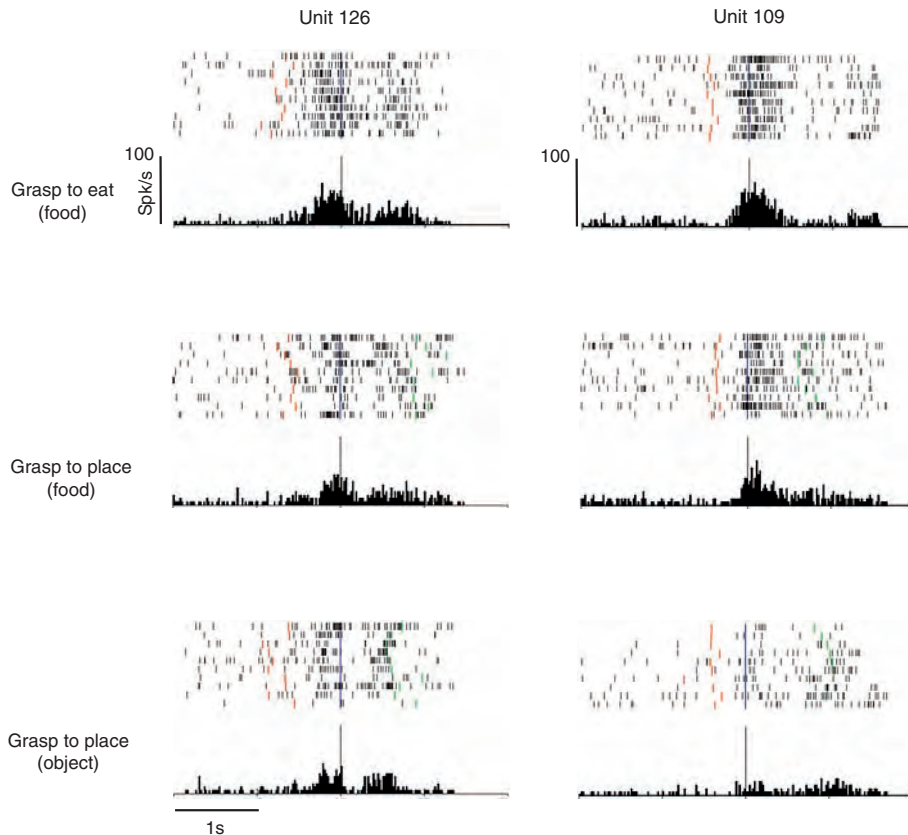


Fig. 4. Visual responses of two IPL mirror neurons during the observation of grasping done by an experimenter. Unit 126 illustrates the case of a neuron selective for grasping to eat whose responsiveness is not influenced by the type of target (food or object) grasped by the experimenter. Unit 109 shows the responses of another grasping-to-eat neuron. As for Unit 126, the visual selectivity remains the same regardless of the object grasped by the experimenter, but the intensity of the discharge is modulated by the type of grasped stimulus.

Table 1. IPL neurons tested during execution and observation of grasping to eat and grasping to place.

(A) Number of neurons tested during the execution of the motor task.			Total
Influenced by the final goal		Not influenced by the final goal	
Eating > Placing	Placing > Eating	Eating = Placing	
77 (72.6%)	29 (27.4%)		
106 (64.2%)		59 (35.8%)	165 (100%)
(B) Number of neurons tested during the observation of the motor task done by the experimenter.			Total
Influenced by the final goal		Not influenced by the final goal	
Eating > Placing	Placing > Eating	Eating = Placing	
23 (74.2%)	8 (25.8%)		
31 (75.6%)		10 (24.4%)	41 (100%)

ing to the final goal of the action, this neuronal property allows the monkey to predict the goal of the observed action and, thus, to “read” the intention of the acting individual.

This mechanism for understanding intention appears to be rather simple. Depending on which motor chain is activated, the observer is going to have an internal representation of what, most likely, the action agent is going to do. What is more complex is to specify how the selection of a particular chain occurs. After all, what the observer sees is just a hand grasping a piece of food or an object.

Various factors may determine this selection. The first is the context in which the action is executed. In our experiment, the presence or absence of the container clearly indicated the action that the experimenter was intending to do. The second factor is the type of object that the experimenter grasped. Typically, food is grasped to be eaten, whereas this is, of course, not the case for solids. Thus, because of this food-action association, the observation of a motor act directed toward food more likely triggers neurons coding grasping to eat than neurons coding grasping for other purposes. This association is, obviously, not mandatory. If it were, the vision of a piece of food would always trigger the chain of motor acts that code eating.

The two factors (context and object type) were found to interact. In several instances (e.g., Unit 109, Fig. 4), we observed that neurons coding grasping to eat also discharged, although weakly, during grasping to place when the object to be placed was food but not when it was a solid. It was as if the eating chain were activated, although slightly, by food despite the presence of contextual cues indicating that placing was the most likely action. A few neurons, instead of showing an intermediate discharge when the nature of the stimulus (food) and context conflicted, decreased their discharge with time when the same action was repeated. It was as if the activity of the placing chain progressively inhibited the activity of neurons of the eating chain (fig. S3).

It is outside the aim of the present study to describe the effects of these factors and their reciprocal influence on the selection of different chains of motor acts in detail. It is important to note, however, that the rich connections of the IPL (24, 25) with areas coding biological actions [areas of the superior temporal sulcus (26, 27)] and object semantics [inferotemporal lobe (28, 29)] render this cortical region particularly suitable for the selection of the coded motor acts according to external contingencies.

Understanding “other minds” constitutes a special domain of cognition (30). Brain imaging studies suggest that several areas might be involved in this function (31–33). Given the complexity of the problem, it would be naïve to claim that the mechanism described in the present study is the sole mechanism underlying mind reading, yet the present data show a neural mechanism through which a basic aspect of understanding intention may be solved. Furthermore, they represent an example of how action and cognition are linked with one another and how the refinement of the motor organization may determine the emergence of complex cognitive functions.

References and Notes

- V. B. Mountcastle, J. C. Lynch, A. Georgopoulos, H. Sakata, C. Acuna, *J. Neurophysiol.* **38**, 871 (1975).
- J. Hyvärinen, *Physiol. Rev.* **62**, 1060 (1982).
- H. Sakata, M. Taira, *Curr. Opin. Neurobiol.* **4**, 847 (1994).
- C. L. Colby, in *Information Integration in Perception and Communication: Attention and Performance*, Vol. XIV, T. Inui, J. L. McClelland, Eds. (MIT Press, Cambridge, MA, 1996), pp. 157–178.
- R. A. Andersen, L. H. Snyder, D. C. Bradley, J. Xing, *Annu. Rev. Neurosci.* **20**, 303 (1997).
- A. Murata, V. Gallese, G. Luppino, M. Kaseda, H. Sakata, *J. Neurophysiol.* **83**, 2580 (2000).
- H. Yokochi, M. Tanaka, M. Kumashiro, A. Iriki, *Somatosens. Mot. Res.* **20**, 115 (2003).
- P. F. Ferrari *et al.*, *Soc. Neurosci. Abs.* **919.7** (2003).
- L. Fogassi, V. Gallese, L. Fadiga, G. Rizzolatti, *Soc. Neurosci. Abs.* **257.5** (1998).
- V. Gallese, L. Fadiga, L. Fogassi, G. Rizzolatti, in *Common Mechanisms in Perception and Action: Attention and Performance*, Vol. XIX, W. Prinz, B. Hommel, Eds. (Oxford Univ. Press, Oxford, 2002), pp. 334–355.
- M. Tanaka, H. Yokochi, A. Iriki, *Soc. Neurosci. Abs.* **82.14** (2004).
- Materials and methods are available as supporting material on Science Online.
- M. Jeannerod, *The Neural and Behavioural Organization of Goal-Directed Movements* (Clarendon Press, Oxford, 1988).
- A. R. Luria, *The Working Brain* (Penguin, London, 1973).
- The chain organization of motor acts discussed here is conceptually different from the sequences of actions previously studied by various authors (34). In our case, various motor acts are naturally chained together to form a single goal-directed action. In studies of sequence learning, motor actions are arbitrarily arranged in an appropriate temporal order.
- P. F. Ferrari *et al.*, unpublished observations.
- M. Petrides, D. N. Pandya, *J. Comp. Neurol.* **228**, 105 (1984).
- M. Matelli, R. Camarda, M. Glickstein, G. Rizzolatti, *J. Comp. Neurol.* **251**, 281 (1986).
- G. Rizzolatti, G. Luppino, M. Matelli, *Electroencephalogr. Clin. Neurophysiol.* **106**, 283 (1998).
- G. Rizzolatti, C. Scandolara, M. Matelli, M. Gentilucci, *Behav. Brain Res.* **2**, 125 (1981).
- G. di Pellegrino *et al.*, *Exp. Brain Res.* **91**, 176 (1992).
- V. Gallese, L. Fadiga, L. Fogassi, G. Rizzolatti, *Brain* **119**, 593 (1996).
- G. Rizzolatti, L. Fogassi, V. Gallese, *Nat. Rev. Neurosci.* **2**, 661 (2001).
- J. S. Baizer, L. G. Ungerleider, R. Desimone, *J. Neurosci.* **11**, 168 (1991).
- G. Luppino *et al.*, *FENS Abstr.* **A025.11** (2004).
- D. I. Perrett *et al.*, *J. Exp. Biol.* **146**, 87 (1989).
- T. Jellema, C. I. Baker, M. W. Oram, D. I. Perrett, in *The Imitative Mind: Development, Evolution, and Brain Bases*, A. N. Melzoff, W. Prinz, Eds. (Cambridge Univ. Press, Cambridge, UK, 2002), pp. 143–162.
- C. G. Gross, C. E. Rocha-Miranda, D. B. Bender, *J. Neurophysiol.* **35**, 96 (1972).
- K. Tanaka, *Curr. Opin. Neurobiol.* **7**, 523 (1997).
- R. Saxe, S. Carey, N. Kanwisher, *Annu. Rev. Psychol.* **55**, 87 (2004).

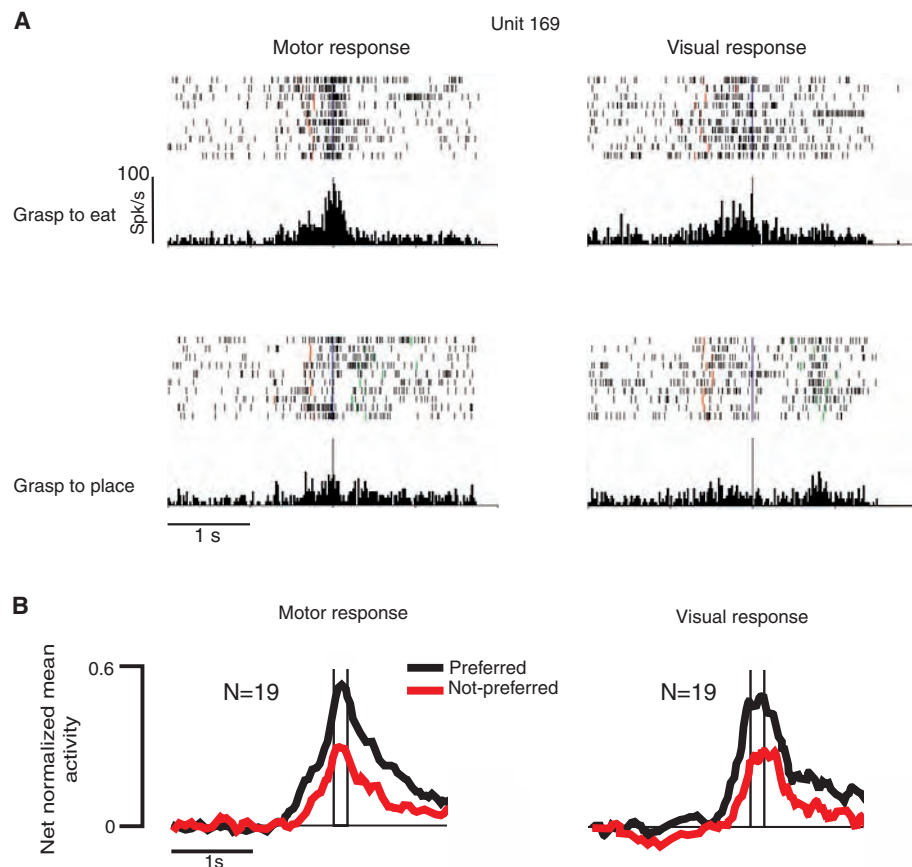


Fig. 5. (A) Congruence between the visual and the motor response of a mirror neuron. Unit 169 has a stronger discharge during grasping to eat than during grasping to place, both when the action is executed and when it is observed. Conventions as in Fig. 1. (B) Population-averaged responses during motor and visual tasks (12).

31. S. J. Blakemore, J. Decety, *Nat. Rev. Neurosci.* **2**, 561 (2001).
 32. U. Frith, C. D. Frith, *Philos. Trans. R. Soc. London Ser. B* **358**, 459 (2003).
 33. H. L. Gallagher, C. D. Frith, *Trends Cogn. Sci.* **7**, 77 (2003).
 34. J. Tanji, *Ann. Rev. Neurosci.* **24**, 631 (2001).
 35. This study was supported by European Union (EU) Contract QLG3-CT-2002-00746, Mirror; EU Contract IST-

2000-29689, Artesimit; Cofin 2002; and FIRB RBNE015ZB4. We thank L. Bonini, C. Maiolini, and E. Monfardini for help in data recording.

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5722/662/DC1

Materials and Methods
 Figs. S1 to S3
 References

7 October 2004; accepted 9 February 2005
 10.1126/science.1106138

Amplification of Acetylcholine-Binding Catenanes from Dynamic Combinatorial Libraries

Ruby T. S. Lam, Ana Belenguer, Sarah L. Roberts,
 Christoph Naumann, Thibaut Jarrosson,
 Sijbren Otto, Jeremy K. M. Sanders*

Directed chemical synthesis can produce a vast range of molecular structures, but the intended product must be known at the outset. In contrast, evolution in nature can lead to efficient receptors and catalysts whose structures defy prediction. To access such unpredictable structures, we prepared dynamic combinatorial libraries in which reversibly binding building blocks assemble around a receptor target. We selected for an acetylcholine receptor by adding the neurotransmitter to solutions of dipeptide hydrazones [proline-phenylalanine or proline-(cyclohexyl)alanine], which reversibly combine through hydrazone linkages. At thermodynamic equilibrium, the dominant receptor structure was an elaborate [2]-catenane consisting of two interlocked macrocyclic trimers. This complex receptor with a 100 nM affinity for acetylcholine could be isolated on a preparative scale in 67% yield.

Dynamic combinatorial chemistry is a powerful approach for the preparation of unpredictable receptors for recognition and catalysis (1–7). Its key feature is the dynamic combinatorial library, in which each library member is assembled from building blocks that bond reversibly to one another under the reaction conditions. As a result of this reversibility, all library members can interconvert to give a distribution that is under thermodynamic control. Addition of a guest molecule, or template, that can selectively recognize one receptor in the library will serve to increase the concentration of that host at the expense of unsuccessful structures in the library. The successful host is then isolated and identified. Such systems display feedback amplification of molecules with the desired properties in a manner that is reminiscent of the mammalian immune system. Typically, however, this approach has produced relatively simple structures such as macrocycles or topologically linear species (8).

As a neurotransmitter, acetylcholine (ACh) is an attractive target for studies of molecular

recognition. We report here that ACh acts as a template to amplify [2]-catenanes from small peptide-hydrazone building blocks whose

properties and chemistry we have previously described (9). In solution, six of these building blocks assemble around ACh into a remarkably complex receptor structure comprising two linked 42-membered rings.

Library formation was initiated by addition of trifluoroacetic acid (43 equivalents) to a room-temperature 20 mM solution of the peptide building block *p*PFm (1) in 95:5 (v/v) chloroform/dimethyl sulfoxide (DMSO) (Fig. 1). High-performance liquid chromatography (HPLC) analysis showed that linear intermediates were formed on a time scale of minutes, converting over the next 60 min into a series of simple cyclic oligomers up to at least the cyclic hexamer. In the absence of added template, the library reached equilibrium in 3 days. The main components of this library were identified by mass spectrometry (MS) as the cyclic dimer (2, $MH^+ = 781.4$), trimer (3, $MH^+ = 1172.5$), and tetramer (4, $MH^+ = 1562.8$), with small amounts of pentamer (5, $MH^+ = 1952.9$), hexamer (6, $MH^+ = 2342.0$), and traces of higher oligomers.

Next, we added of 200 mM ACh chloride to the solution to serve as a binding tem-

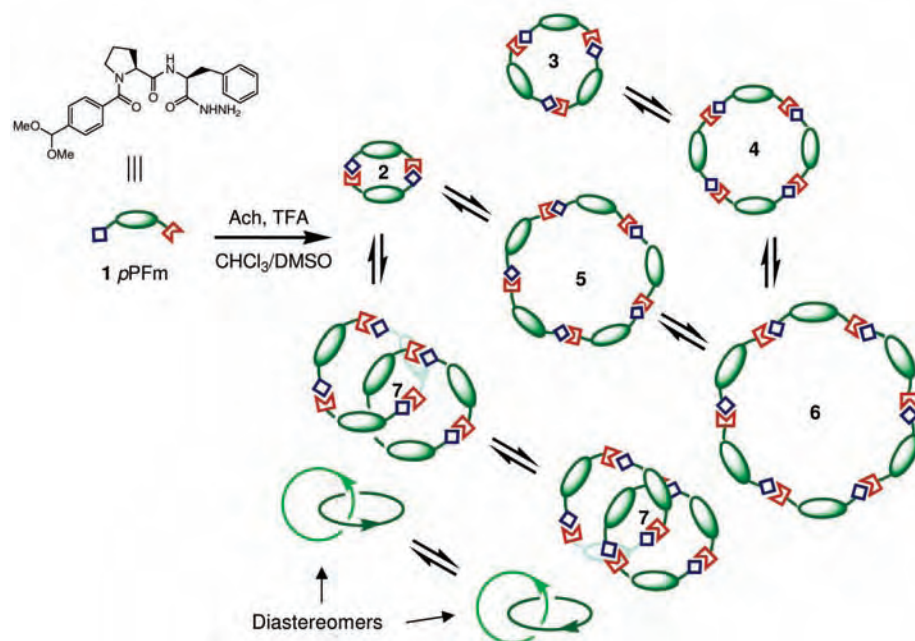


Fig. 1. A *p*PFm dynamic combinatorial library.

Department of Chemistry, University of Cambridge,
 Lensfield Road, Cambridge CB2 1EW, UK.

*To whom correspondence should be addressed.
 E-mail: jkms@cam.ac.uk

plate. We found that the cyclic dimer **2** was amplified initially, becoming the major product of the library (Fig. 2). However, this was a ki-

netically rather than a thermodynamically favored product, and within an hour a new product, **7**, isomeric with the cyclic hexamer ($MH^+ =$

2342.0), was formed at the expense of all the other material in the library. This compound was not observed in the absence of ACh. Both **6** and **7** display a doubly charged molecular ion at mass/charge ratio (m/z) 1172 (Fig. 3, A and B), but they show quite different electrospray MS/MS fragmentation behavior. The simple hexamer, **6**, fragments sequentially to pentamer and smaller components (Fig. 3D), whereas the new templated product, **7**, fragments directly to trimer (Fig. 3C). This fragmentation behavior is typical of a [2]-catenane (*10*) and indicates that the product consists of two identical interlocked cyclic trimers. Because each ring is chiral, there are two possible diastereomers of the [2]-catenane (Fig. 1); the nuclear magnetic resonance (NMR) evidence presented below confirms the gross structure proposed and indicates that only one of the two possible diastereomers is amplified. The 1:1 catenane:ACh complex can also be observed as the singly charged species at m/z 2488.0 by electrospray MS.

The [2]-catenane continued to accumulate over a period of many days, eventually reaching 70% of the material in the library as judged by HPLC; it must have been formed via a pathway that is inherently improbable, and therefore its evolution and accumulation were slow, but as the product was formed it was stabilized by the template. The isolated yield of [2]-catenane was 67%. Because the HPLC detection limit for [2]-catenane is well below 1%, this corresponds to an amplification of more than three orders of magnitude.

Fig. 2. HPLC traces (absorption at 290 nm) of the ρ PfM libraries. (A) In the presence of ACh, a series of cyclic and linear intermediates are observed 2 min after initiation of library formation. (B) [2]-Catenane (**7**) is first observed within 25 min. (C) Cyclic dimer **2** is substantially amplified in the first 3 days. (D) The yield of ACh receptor **7** is optimized after 44 days of equilibration. (E) In the absence of ACh, the ρ PfM library reaches equilibrium in 3 days. HPLC analyses were performed with a reversed-phase Symmetry C18 column (250 mm by 4.6 mm, particle size 5 μ m) using gradient elution of tetrahydrofuran and water at 45°C.

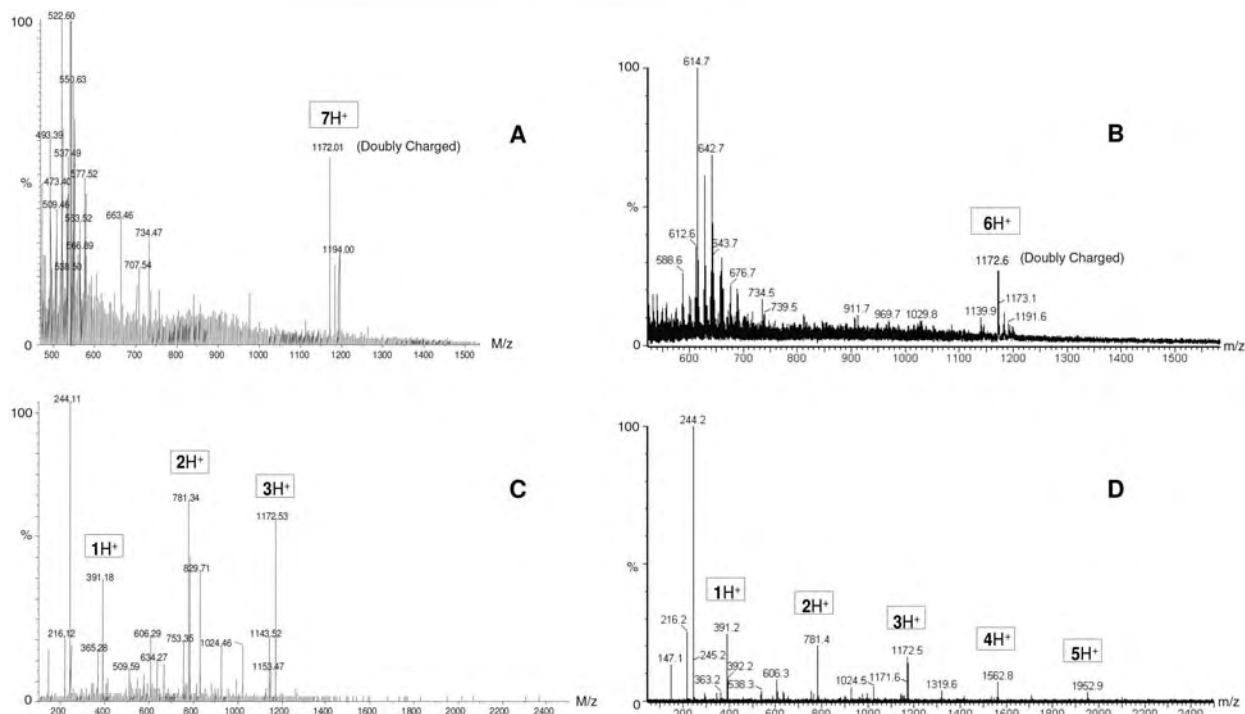
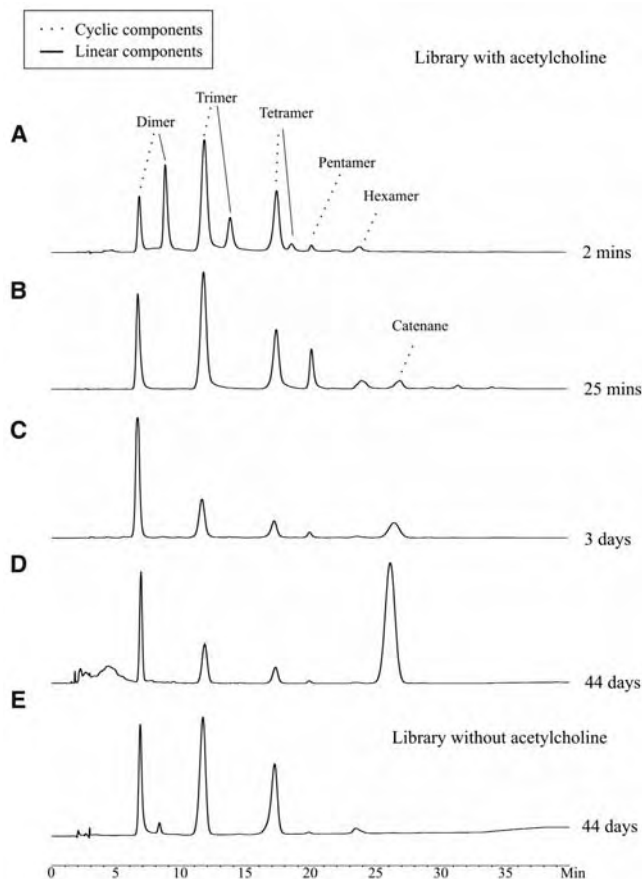


Fig. 3. (A) Electrospray MS spectrum of [2]-catenane ($7H^+$ at 1172.0) shows the same molecular ion mass as that of (B) hexamer ($6H^+$ at 1172.6). (C) Electrospray MS/MS of $7H^+$ displays the direct fragmentation

of [2]-catenane to trimer, differentiating it from hexamer, which shows (D) sequential fragmentation of $6H^+$ into pentamer, tetramer, trimer, dimer, and monomer.

The ^1H NMR spectrum of the [2]-catenane in 95:5 $\text{CDCl}_3/\text{DMSO}-d_6$ displays large clumps of broad signals that indicate the presence of several slowly interconverting conformers (Fig. 4B). Upon addition of ACh or related ligands, the spectrum sharpens and reveals the presence of three inequivalent sets of building blocks in slow exchange (Fig. 4A). Thus, in the complex with ACh, the two rings are related by symmetry but the rotation of one ring with respect to the other is slow on the NMR time scale. The relative simplicity of the spectrum of the catenane-ACh complex indicates (i) that only one diastereomer is present, and (ii) that ACh selects only one conformation of the [2]-catenane from the many possibilities. We previously observed the latter behavior in a much simpler ACh-binding peptide-hydrazone macrocycle (11).

^1H and ^{13}C NMR spectra of the complex in CD_2Cl_2 show the same symmetry and very similar chemical shifts, but binding is weaker in this solvent and allows exchange signals between free and bound ACh to be observed in nuclear Overhauser effect (NOE) spectra. Bound ACh proton signals are shifted upfield by 0.4 to 1.7 parts per million (ppm) (Fig. 4), and both sets of methylene protons become diastereotopic in the chiral environment, as expected (11). Intramolecular NOEs within the [2]-catenane and within the bound ACh resonances confirm the assignments shown

in Fig. 4, but no NOEs are observed between host and guest.

The amplification of a receptor containing six building blocks from a library that contains substantial concentrations of dimers and trimers is statistically highly unfavorable; it demands a strongly selective binding of ACh for the [2]-catenane relative to the smaller and simpler macrocycles (12). This conclusion was confirmed by isothermal titration microcalorimetry (ITC): Strongly exothermic binding was observed (binding constant $K = 1.4 \times 10^7 \text{ M}^{-1}$, enthalpy change $\Delta H = -22.9 \text{ kJ mol}^{-1}$, entropy change $T\Delta S = +17.8 \text{ kJ mol}^{-1}$) between [2]-catenane and ACh in 95:5 $\text{CHCl}_3/\text{DMSO}$. Binding constants of ACh to the trimer **3** and tetramer **4** were much smaller: $K = 1.5 \times 10^3 \text{ M}^{-1}$ and $K = 5.7 \times 10^3 \text{ M}^{-1}$, respectively. The dimer is rather insoluble; no binding was observed in a saturated solution (0.06 mM).

Several sets of observations suggest that the trimethylammonium moiety of ACh is the primary recognition site. Choline itself is too insoluble in our solvent system to allow effective templating or ITC experiments, but it is solubilized by the [2]-catenane; the symmetry and chemical shifts of the choline-catenane complex are essentially identical to those of the ACh complex. In a ^1H NMR competition experiment, two sets of equally intense proton signals were observed, which shows that the [2]-catenane binds choline and ACh with simi-

lar affinity. Replacing one methyl group in ACh by ethyl, to give $\text{Me}_2\text{EtN}^+\text{CH}_2\text{CH}_2\text{OCOMe}$, reduces K by a factor of 100 to 10^5 M^{-1} ; this result confirms the importance of the Me_3N^+ group (13). Lengthening the template esterifying chain from acetyl to butyryl has a minimal effect on binding or templating efficiency.

Replacement of the side-chain aromatic group of phenylalanine in **1** by a cyclohexyl group to give the pPCm analog (9) led to a very similar library, from which the analogous [2]-catenane was amplified by ACh. Its yield, NMR spectrum, and binding properties ($K = 3.0 \times 10^6 \text{ M}^{-1}$) were similar to those of the original [2]-catenane; hence, the aromatic side chain is not essential to the binding process.

The molecular recognition processes and the three-dimensional structures underlying the results described here have yet to be elucidated in detail. However, these results indicate that the potential for dynamic combinatorial chemistry is greater than previously envisaged. Given sufficient time, highly unexpected structures can be formed by templating in a single flask and isolated on a preparative scale. There are many reports of carefully designed [2]-catenane syntheses (14–16) and occasional reports of accidental [2]-catenane synthesis (17), but the results described here demonstrate the creation of a [2]-catenane that is, by virtue of its mode of synthesis, functionally highly effective as a receptor.

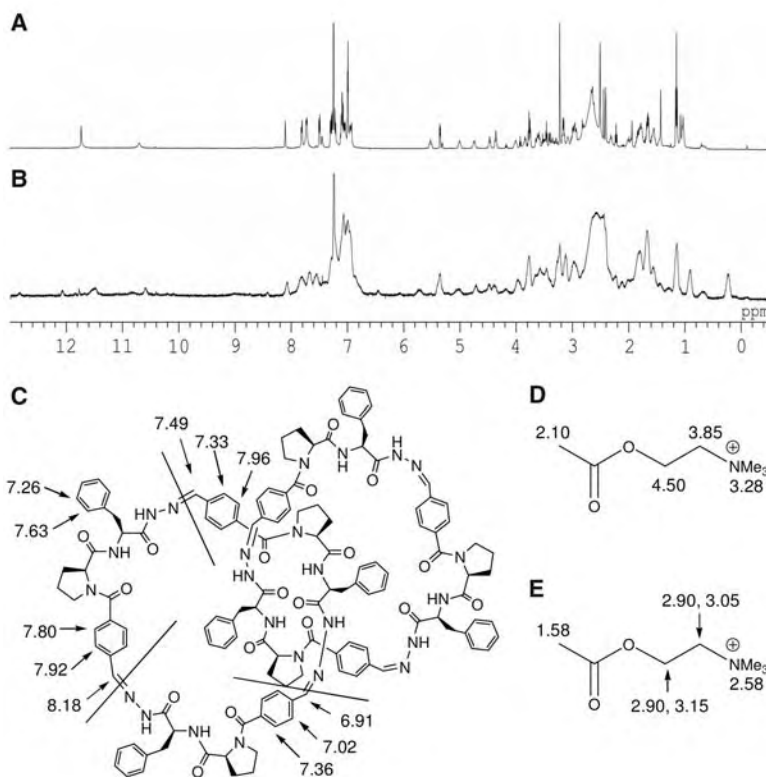


Fig. 4. (A and B) ^1H NMR spectra (13.0 to 0 ppm) of [2]-catenane in $\text{CDCl}_3/\text{DMSO}-d_6$ at 300 K in the presence of 2.0 equiv of ACh and 4 μl of D_2O (A) and in the absence of ACh (B). (C) Structure of **7** with ^1H NMR chemical shifts of some key protons in the catenane-ACh complex. Assignments to building blocks are arbitrary. (D and E) Chemical shifts of free ACh (D) and ACh bound to compound **7** (E).

References and Notes

- P. A. Brady, R. P. Bonar-Law, S. J. Rowan, C. J. Suckling, J. K. M. Sanders, *Chem. Commun.* **35**, 2143 (1996).
- J.-M. Lehn, A. V. Eliseev, *Science* **291**, 2331 (2001).
- O. Ramström, J.-M. Lehn, *Nat. Rev. Drug Discov.* **1**, 26 (2002).
- S. Otto, R. L. E. Furlan, J. K. M. Sanders, *Science* **297**, 590 (2002).
- S. J. Rowan, S. J. Cantrill, G. R. L. Cousins, J. K. M. Sanders, J. F. Stoddart, *Angew. Chem. Int. Ed. Engl.* **41**, 898 (2002).
- S. Otto, *Curr. Opin. Drug Discov. Dev.* **6**, 509 (2003).
- B. Brisig, J. K. M. Sanders, S. Otto, *Angew. Chem. Int. Ed. Engl.* **42**, 1270 (2003).
- S. Otto, S. Kubik, *J. Am. Chem. Soc.* **125**, 7804 (2003).
- S. L. Roberts, R. L. E. Furlan, S. Otto, J. K. M. Sanders, *Org. Biomol. Chem.* **1**, 1625 (2003).
- C. Dietrich-Buchecker, J.-P. Sauvage, *Chem. Commun.* **1999**, 615 (1999).
- G. R. L. Cousins, R. L. E. Furlan, Y.-F. Ng, J. E. Redman, J. K. M. Sanders, *Angew. Chem. Int. Ed. Engl.* **40**, 423 (2001).
- P. T. Corbett, S. Otto, J. K. M. Sanders, *Chem. Eur. J.* **10**, 3139 (2004).
- P. Sarri, F. Venturi, F. Cuda, S. Roelens, *J. Org. Chem.* **69**, 3654 (2004).
- S. A. Vignon, J. Wong, H.-R. Tseng, J. F. Stoddart, *Org. Lett.* **6**, 1095 (2004).
- L. Raehm, D. G. Hamilton, J. K. M. Sanders, *Synlett* (no. 11), 1743 (2002).
- A. Godt, *Eur. J. Org. Chem.* **8**, 1639 (2004).
- C. A. Hunter, *J. Am. Chem. Soc.* **114**, 5303 (1992).
- Supported by the UK Biotechnology and Biological Sciences Research Council, the UK Engineering and Physical Sciences Research Council, the Royal Society, GlaxoSmithKline, and the Croucher Foundation.

21 January 2005; accepted 1 March 2005

Published online 10 March 2005;

10.1126/science.1109999

Include this information when citing this paper.

Experimental Verification of Designer Surface Plasmons

Alastair P. Hibbins,* Benjamin R. Evans, J. Roy Sambles

We studied the microwave reflectivity of a structured, near perfectly conducting substrate that was designed to verify the existence of a theoretically proposed new class of surface mode. Measurements of the mode's dispersion curve show that it correctly approaches the predicted asymptotic frequency; the curve also agrees well with that derived from a computer simulation. Modeling of the field distribution on resonance provides evidence of strong localization of the electric field at the interface and substantial power flow along the interface, thus verifying the surface plasmon-like nature of the mode.

Conductors support electromagnetic surface waves (1, 2) over a broad range of frequencies, ranging from dc and radio frequencies up to the visible. These modes propagate along the interface with a dielectric, and they comprise an electromagnetic field coupled to the oscillations of conduction electrons. At visible wavelengths, the mode that travels along a metal surface is called a surface plasmon (SP). It is characterized by strongly enhanced fields at the interface, which decay exponentially with distance [on the order of 10 nm in the metal and ~ 100 nm in the dielectric (3–5)]. However, in the microwave regime, metals can often be treated as near-perfect conductors. The fields associated with surface waves at these frequencies penetrate ~ 1 μm into the metal but extend many hundreds of wavelengths into the dielectric above (Fig. 1A). In other words, this mode is not bound at the interface and is simply a surface current.

It was recently predicted (6) that even in the perfectly conducting limit, the same metal, textured with subwavelength holes, can support strongly localized SP-like waves. This is because the holes may allow some of the field to penetrate into the substrate, hence changing the field-matching situation at the surface (Fig. 1B). The holes in the metal act as waveguides and therefore have a cutoff frequency below which no propagating modes are allowed. Hence, below cutoff, only evanescent fields exist on the metal side of the interface, and it is exactly this field characteristic that is required for a surface mode. The cutoff frequency of the modes in the waveguides is equivalent to an effective “surface plasma” frequency of the structured surface, which for holes of square cross section ($a \times a$) is given by

$$v_{\text{cutoff}} = \frac{c}{2a\sqrt{\epsilon_h\mu_h}} \quad (1)$$

School of Physics, University of Exeter, Stocker Road, Exeter EX4 4QL, UK.

*To whom correspondence should be addressed. E-mail: a.p.hibbins@exeter.ac.uk

where c is the velocity of light in vacuum and ϵ_h and μ_h are the relative permittivity and permeability of the material filling the holes, respectively. Hence, this frequency can be engineered to occur at almost any frequency below the natural surface plasma frequency of a metal [which is usually in the ultraviolet range (4)].

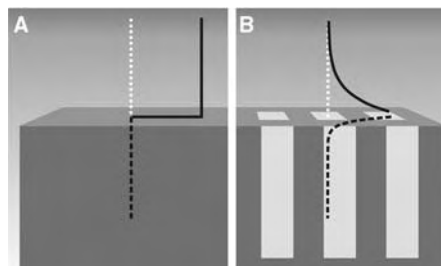


Fig. 1. Schematic representation of electric fields associated with a mode propagating along the surface of a metal. At microwave frequencies, the metal is almost perfectly conducting. (A) The field is almost completely excluded from the substrate but extends for many hundreds of wavelengths into the dielectric region above. The mode is essentially a surface current. (B) An effective penetration depth is achieved by perforating the substrate with an array of subwavelength holes. The holes allow the fields to decay exponentially into the structure, and they closely resemble those of a surface plasmon propagating on metals in the visible regime.

Fig. 3. The dispersion of the surface mode on the air-filled sample. Radiation is TM-polarized and incident in the xz plane ($\varphi = 0^\circ$). The frequency of the resonance is derived from experimental (\times) and modeled (solid line) data sets. The shaded region corresponds to momentum space within which surface modes may not be directly coupled to, as they are beyond the light line. The dashed line represents the first-order diffracted light lines centered on $k_x = \pm 2\pi/(2d) = \pm 330 \text{ m}^{-1}$ associated with the array of cylindrical rods. The dot-dashed line similarly corresponds to first-order diffraction from the array of brass tubes. (Inset) TM-polarized reflectivity spectrum obtained at $\theta = \sim 14^\circ$, illustrating the resonant surface mode at ~ 12.3 GHz.

We provide experimental evidence of the resonance of a surface mode propagating across the surface of a near perfectly conducting substrate perforated with holes. By carefully choosing the size and periodicity of the holes, the effective “surface plasma” frequency has been lowered to the microwave regime. We have observed resonant absorption at different angles of incidence (θ) and have thereby been able to plot much of the dispersion curve. The results agree exceptionally well with those determined using a finite element method (FEM) model (7), and we are able to witness the mode's asymptotic approach to the predicted effective “surface plasma” frequency limit. The model is also used to calculate and illustrate the field solutions on resonance, and hence to verify the SP-like nature of the mode.

Our sample is a 300 mm \times 300 mm array of hollow, square-ended brass tubes of length 45 mm, side length $d = 9.525$ mm, and inner dimension $a = 6.960$ mm. From Eq. 1, $v_{\text{cutoff}} = 21.54$ GHz. The tubes are carefully arranged, square face down, on a flat brass plate and tightly clamped together. Note that the plate supporting the tubes will have no substantial effect on the surface mode, because we are only interested in frequencies below v_{cutoff} . A modification of the originally proposed structure (6) is the addition of a periodic array of cylindrical rods (radius $r = 1.0$ mm) with pitch $2d$ positioned on the surface of the array of tubes (Fig. 2). These allow for control of the strength of diffractive coupling to the

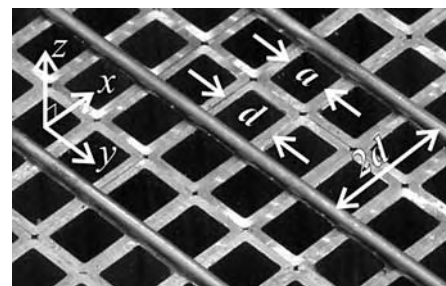
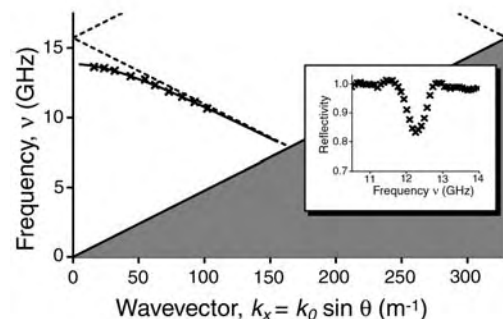


Fig. 2. Photograph of the experimental sample showing the square brass tubes ($d = 9.525$ mm, $a = 6.960$ mm) and brass cylindrical rods (radius $r = 1.0$ mm). The coordinate system is also shown.



mode, which would otherwise be very weak. In addition, relying on the d periodicity alone would only allow for observation of the surface mode either close to v_{cutoff} or at very high values of θ . The former could result in confusion with the onset of propagating waveguide modes, whereas the latter is not experimentally favorable. However, the first-order “ $2d$ ” diffracted light lines associated with the rods will cross the frequency axis at ~ 15.7 GHz. Because a band gap in the dispersion of the surface mode will be established at this crossing (normal incidence), the mode’s resonant frequency will be reduced below this point. For this reason, the reflectivity features associated with the surface mode cannot be confused with the onset of diffraction.

Our initial set of reflectivity measurements was recorded at a series of fixed values of θ with collimated millimeter-wave radiation ($10 \text{ GHz} \leq \nu \leq 15 \text{ GHz}$) incident in the xz plane ($\varphi = 0^\circ$). By positioning the transmitting horn antenna at the focus of a mirror with radius of curvature 4 m, we reduced the amount of beam spread in the system. The specularly reflected signal was then detected by a second horn antenna, and the orientations of both horn antennae were set to pass only transverse magnetic (TM)-polarized radiation. The inset of Fig. 3 illustrates a typical experimental reflectivity spectrum recorded at $\theta = \sim 14^\circ$. The main part of Fig. 3 shows the distribution of the resonant mode positions derived from the experimental and

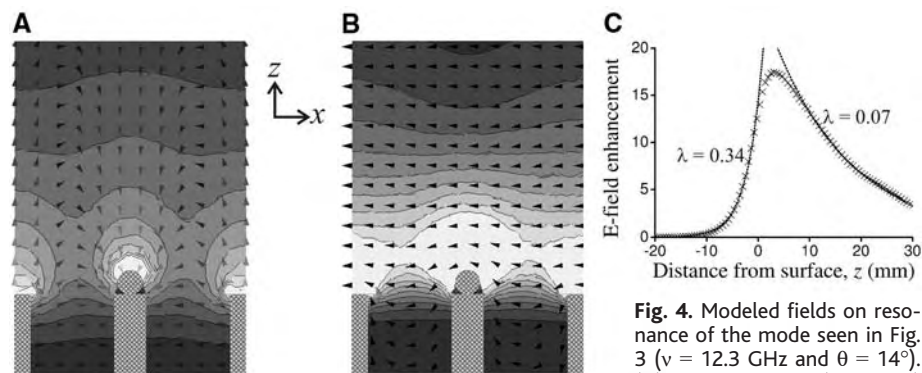
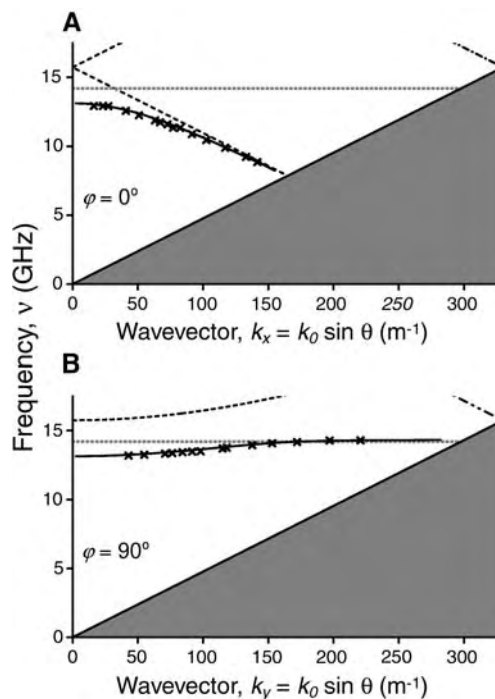


Fig. 4. Modeled fields on resonance of the mode seen in Fig. 3 ($\nu = 12.3 \text{ GHz}$ and $\theta = 14^\circ$). (A) Time-averaged (grayscale)

and instantaneous (arrowheads) electric field strengths, where the latter is plotted at a phase corresponding to maximum enhancement. (B) Poynting vector direction (arrowheads) and magnitude (grayscale). White areas correspond to enhancement in electric field strength by a factor of ≥ 40 and of power flow by a factor of 300. (C) Time-averaged electric field strength calculated for either side of the sample surface ($z = 0$) at the center of a tube waveguide (i.e., at a perpendicular distance of $a/2$ from the inside walls). The decays on either side of the peak have been fitted to single exponentials (λ , decay constant).

Fig. 5. The dispersion of the surface mode on the wax-filled brass tube array. Radiation is TM-polarized and incident in (A) the xz plane ($\varphi = 0^\circ$) and (B) the yz plane ($\varphi = 90^\circ$). The frequency of the resonance is derived from experimental (\times) and modeled (solid line) data sets. The shaded region corresponds to momentum space within which surface modes may not be directly coupled to, as they are beyond the light line. The dashed lines represent the first-order diffracted light lines associated with the array of cylindrical rods. The dot-dashed lines similarly correspond to first-order diffraction from the square array of brass tubes. Horizontal dotted lines correspond to v_{cutoff}



modeled reflectivity spectra as a function of in-plane momentum ($k_x = k_0 \sin \theta$) and frequency. Although there is a small spread in the experimental data points (associated with the difficulty in making accurate measurements of θ), the experiment is in excellent agreement with the modeled dispersion curve.

Further verification of the SP-like characteristics of the mode is provided by examining the modeled field distributions at the resonant frequency. For example, consider the resonance of the mode shown in the inset of Fig. 3 ($\nu = 12.3 \text{ GHz}$). The time-averaged electric field strength is shown in Fig. 4A with the instantaneous electric field vector distribution (plotted at a time in phase corresponding to maximum enhancement) superimposed. Regions of strongest field are clearly located directly above the metal regions of the surface, and the field-line loops are highly reminiscent of the SP fields on metals (4, 5). The data in Fig. 4C further strengthen the SP argument, showing the exponentially decaying electric field strength away from the surface. Note how the exponential decay constant into the effective medium is about one-fifth that into the dielectric half-space above. The Poynting vector plot (Fig. 4B) provides further verification of the nature of the mode: The region of greatest power flow is strongly enhanced just above the surface, flowing parallel to the average surface plane (as would be expected for a SP).

To summarize so far, we have associated a measured dip in reflectance with the proposed surface modes (6) and have shown that the field profiles on resonance are strongly reminiscent of SPs. However, the dispersion of the mode does not asymptotically approach v_{cutoff} ; instead, it is perturbed by the SP band gap at normal incidence. (Note that only the lower energy band of the gap may be coupled to; the upper band mode would correspond to charge located in the voids.) We now show that the asymptotic approach to v_{cutoff} can be achieved by rotating the plane of incidence by 90° (i.e., the yz plane, $\varphi = 90^\circ$) and filling the tubes with a dielectric.

First consider the dispersion of the mode supported by a structure in the original $\varphi = 0^\circ$ geometry, but this time filling the tubes with wax [$\epsilon_{\text{wax}} = 2.3$ (8)] (Fig. 5A). Remember that the mode observed in the $\varphi = 0^\circ$ geometry is coupled via the evanescent fields of the +1 “ $2d$ ” diffracted order (due to the rods). Although there is a small reduction of the resonant frequency close to normal incidence, the dispersion looks similar to that of the air-filled sample because the perturbation is still dominated by the effect of the band gap. Once again, only the lower energy band can be coupled to, and therefore no asymptotic approach to v_{cutoff} is seen. Now consider how the band diagram changes by rotating the plane of incidence by 90° .

Clearly, at normal incidence, the resonant frequency of this mode must remain unchanged, as must the frequency at which the diffracted order begins to propagate. In the rotated geometry ($\varphi = 90^\circ$), diffraction due to the rods will be in a plane orthogonal to the plane of incidence, and the frequency associated with the onset of diffraction will increase hyperbolically with wave vector ($k_y = k_0 \sin \theta$). Our measurements and modeling show (Fig. 5B) that close to normal incidence, the dispersion of the mode exhibits the expected dependence. The mode asymptotically approaches a frequency $\sim v_{\text{cutoff}}$ for larger values of k_y . A small difference between

v_{cutoff} and the observed limit is not surprising, as Eq. 1 does not take into account the finite length of the tubes. The results from the wax-filled sample thus fully confirm the expected behavior.

The idea of “designer” surface modes proposed by Pendry *et al.* (6) promises the ability to engineer a SP at almost any frequency. Our results verify the propagation of SP-like modes on structured, near perfectly conducting substrates. These new “metamaterials” offer the ability of applying near-field, surface plasmon-induced concepts, which have been well studied in the visible regime, to the microwave domain.

All-Optical Switching in Rubidium Vapor

Andrew M. C. Daves, Lucas Illing, Susan M. Clark,
Daniel J. Gauthier*

We report on an all-optical switch that operates at low light levels. It consists of laser beams counterpropagating through a warm rubidium vapor that induce an off-axis optical pattern. A switching laser beam causes this pattern to rotate even when the power in the switching beam is much lower than the power in the pattern. The observed switching energy density is very low, suggesting that the switch might operate at the single-photon level with system optimization. This approach opens the possibility of realizing a single-photon switch for quantum information networks and for improving transparent optical telecommunication networks.

An important component of high-speed optical communication networks is an all-optical switch, where an incoming “switching” beam redirects other beams through light-by-light scattering in a nonlinear optical material (1, 2). For quantum information networks, it is important to develop optical switches that are actuated by a single photon (3). Unfortunately, because the nonlinear optical interaction strength of most materials is so small, achieving single-photon switching is difficult. This problem appears to be solved through modern quantum-interference methods, in which the nonlinear interaction strength can be increased by many orders of magnitude (3–10). It is also important to develop all-optical switches where the output beam is controlled by a weaker switching beam, so they can be used as cascaded classical or quantum computational elements (11). Current switches, in contrast, tend to control a weak beam with a strong one.

In this Report, we describe an all-optical switch that combines the extreme sensitivity of instability-generated transverse opti-

cal patterns to tiny perturbations (12–16) with quantum-interference methods (3–10). A transverse optical pattern is the spatial structure of the electromagnetic field in the plane perpendicular to the propagation direction. As an example, the transverse optical pattern corresponding to two beams of light is a pair of spots. We control such a pattern with a beam whose power is a factor of 6500 times smaller than the power contained in the pattern itself, verifying that the switch is cascable. Also, the switch is actuated with as few as 2700 photons and thus operates in the low-light-level regime. A measured switching energy density $E \sim 3 \times 10^{-3}$ photons/($\lambda^2/2\pi$), where $\lambda = 780$ nm is the wavelength of the switching beam, suggests that the switch might operate at the single-photon level with system optimization such as changing the pump-beam size or vapor cell geometry (17).

Our experimental setup consists of a weak switching beam that controls the direction of laser beams emerging from a warm laser-pumped rubidium vapor. Two pump laser beams counterpropagate through the vapor and induce an instability that generates new beams of light (i.e., a transverse optical pattern) when the power of the pump beams is above a critical level (17), as shown schematically in Fig. 1. The instability arises from

References and Notes

1. H. M. Barlow, A. L. Cullen, *Proc. IEE* **10**, 329 (1953).
2. R. Collin, *Field Theory of Guided Waves* (Wiley, New York, ed. 2, 1990), chap. 7.
3. R. H. Ritchie, *Phys. Rev.* **106**, 874 (1957).
4. H. Raether, *Surface Plasmons* (Springer-Verlag, Berlin, 1988), chap. 2.
5. W. L. Barnes, A. Dereux, T. W. Ebbesen, *Nature* **424**, 824 (2003).
6. J. B. Pendry, L. Martín-Moreno, F. J. García-Vidal, *Science* **305**, 847 (2004); published online 8 July 2004 (10.1126/science.1098999).
7. FEM computer modeling was done with the use of HFSS (Ansoft Corporation, Pittsburgh, PA).
8. A. P. Hibbins, J. R. Sambles, C. R. Lawrence, *J. Appl. Phys.* **87**, 2677 (2000).

22 December 2004; accepted 8 March 2005
10.1126/science.1109043

mirrorless parametric self-oscillation (17–26) due to the strong nonlinear coupling between the laser beams and atoms. Mirrorless self-oscillation occurs when the parametric gain due to nonlinear wave-mixing processes becomes infinite. Under this condition, infinitesimal fluctuations in the electromagnetic field strength trigger the generation of new beams of light. The threshold for this instability is lowest (and the parametric gain enhanced) when the frequency of the pump beams is set near the $^{87}\text{Rb } ^5\text{S}_{1/2} \leftrightarrow ^5\text{P}_{3/2}$ resonance (780-nm transition wavelength). The setup is extremely simple in comparison to most other low-light-level all-optical switching methods (7, 8), and the spectral characteristics of the switching and output light match well with recently demonstrated single-photon sources and storage media (27, 28).

For a perfectly symmetric experimental setup, the instability-generated light (referred to henceforth as “output” light) is emitted both forward and backward along cones centered on the pump beams, as shown in Fig. 1A. The angle between the pump-beam axis and the cone is on the order of ~ 5 mrad and is determined by competition between two different nonlinear optical processes: backward four-wave mixing in the phase-conjugation geometry and forward four-wave mixing (17, 23, 24). The generated light has a state of polarization that is linear and orthogonal to that of the linearly copolarized pump beams (25); hence, it is easy to separate the output and pump light with the use of polarizing elements. Once separated, the output light propagating in one direction (e.g., the forward direction) can appear as a ring on a measurement screen that is perpendicular to the propagation direction and in the far field (Fig. 1, A and B). This ring is known as a transverse optical pattern (18) and is one of many patterns that occur in a wide variety of nonlinear systems spanning the scientific disciplines (29).

Weak symmetry breaking caused by slight imperfections in the experimental setup reduces the symmetry of the optical pattern and selects its orientation (23). For high pump

Department of Physics, Duke University, Box 90305, Durham, NC 27708, USA.

*To whom correspondence should be addressed.
E-mail: gauthier@phy.duke.edu

powers, the pattern consists of six spots with sixfold symmetry, as shown schematically in Fig. 1C. For lower powers near the instability threshold, only two spots appear in the far field in both directions (forward and backward), as shown schematically in Fig. 1D. The azimuthal angle of the spots (and the corresponding beams) is dictated by the system asymmetry, which can be adjusted by slight misalignment of the pump beams or application of a weak magnetic field. The orientation of the spots is stable for several minutes in the absence of a switching beam.

In our all-optical switch, the direction of the bright output beams (total power P_{out}) is controlled by applying a weak switching laser beam with a state of polarization that is linear and orthogonal to that of the pump beams (Fig. 2). The azimuthal angle of the output beams is extremely sensitive to tiny perturbations because the symmetry breaking of our setup is so small (17). Directing the switching beam along the conical surface (Fig. 2B) causes the output beams to rotate to a new angle, while P_{out} remains essentially unchanged. Typically, the orientation of the output beams rotates to the direction of the

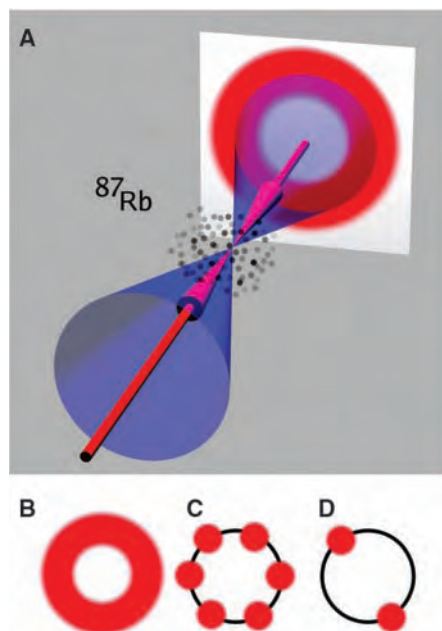


Fig. 1. Generation and symmetry of transverse optical patterns. (A) Two linearly copolarized pump beams (red) counterpropagate through warm ^{87}Rb vapor. A modulational instability generates orthogonally polarized light, which is emitted along cones (blue) and forms a transverse optical pattern (red) on a screen perpendicular to the propagation direction. (B to D) Schematic of the transverse optical pattern for (B) a perfectly symmetric setup, (C) weakly broken symmetry and high pump power, and (D) weakly broken symmetry and low pump power.

switching beam and we find that the pattern is most easily perturbed when the switching beam is injected at azimuthal angles of $\pm 60^\circ$, thereby preserving the sixfold symmetry of the pattern observed for higher pump powers. Notably, the switching beam also controls the output beams in the backward direction.

We now describe the behavior of our switch for two values of the peak power P_s of the switching beam, where the spots rotate by -60° in the presence of a switching beam. In the absence of a switching beam ($P_s = 0$), we observed the pattern shown in Fig. 3A, where $P_{\text{out}} = 1.5 \mu\text{W}$ and the total power emitted from the vapor cell in the forward direction in both states of polarization is $19 \mu\text{W}$. For the higher power switching beam ($P_s = 2.5 \text{ nW}$), we observed complete rotation of the output beams (Fig. 3B), whereas we found that only approximately half the power in the beams rotates to the new azimuthal angle (Fig. 3C) at the lower switching power ($P_s = 230 \text{ pW}$). However, we observed high-contrast switching in both cases.

To quantify the dynamic behavior of the device, we turned the switching beam on

and off and measured the resulting change in the output beams. Figure 4, A and B, show the temporal evolution of the power for $P_s = 2.5 \text{ nW}$ on a coarse time, where it is seen that power variation at each location is out of phase, indicating full redirection of the optical power (movie S1).

At higher temporal resolution (Fig. 4C), weak periodic modulation of the emitted light due to a dynamic instability (25) is apparent. This modulation corresponds to small fluctuations in the power of the pattern, although its orientation is stable. Even in the presence of this modulation, the contrast-to-noise ratio of the switch is at least 30:1 (defined as the change in power between the on and off states: root-mean-square value of the fluctuations). The rise time of the switch is $\tau = 4 \mu\text{s}$, which we believe is largely governed by the rubidium ground-state optical pumping time (23).

Under the conditions shown in Fig. 3B, the total power in the output beams in the forward direction is equal to $P_{\text{out}} = 1.5 \mu\text{W}$, whereas the power of the input switching beam is only $P_s = 2.5 \text{ nW}$. That is, a switching beam controls the behavior of output

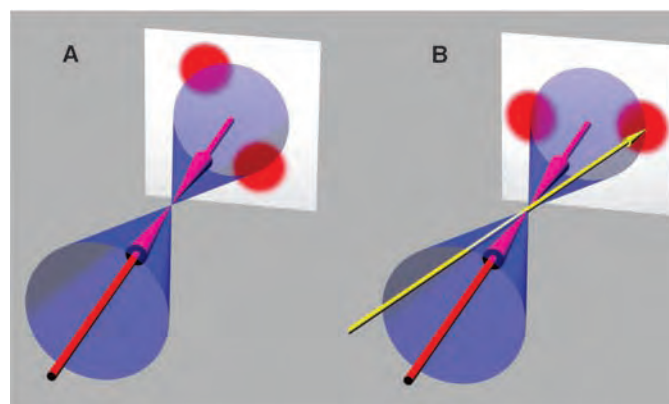


Fig. 2. The two states of the switch. (A) The off state. Weak symmetry breaking results in a two-spot output pattern. (B) The on state. A weak switching beam (yellow), directed along the cone (blue), causes the output pattern to rotate. The state of polarization of the switching beam (yellow) is linear and orthogonal to that of the pump beams (red).

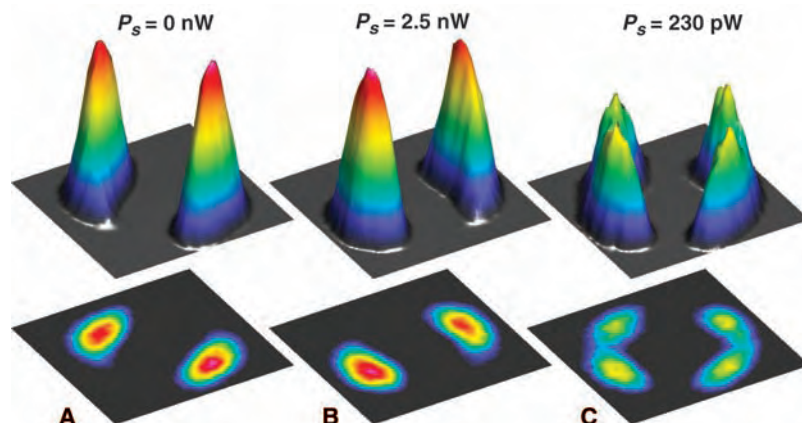


Fig. 3. Low-light-level all-optical switching. The lower panels show a false-color rendition of the detected optical power of the output light (gray, low power; red, high power), and the upper panels are a three-dimensional representation of the same data. (A) The off state with $P_s = 0$. The output light forms a two-spot transverse optical pattern. (B) The on state with $P_s = 2.5 \text{ nW}$. The two-spot output pattern is rotated by -60° . (C) The on state with $P_s = 230 \text{ pW}$. Approximately half the output power is rotated by -60° .

beams that are at least 600 times stronger. Based on the time response of the switch, the number of photons needed to change its state is given by $N_p = \tau P_s / E_p = 40,000$, where $E_p = 2.55 \times 10^{-19}$ J is the photon energy, and the switching energy is equal to $N_p E_p = 10$ fJ. Optical switches that operate with similarly low energies have been proposed previously, but their application to few-photon switching has not been discussed (12, 14, 15).

Another metric to characterize low-light-level switches is the energy density E , given in units of photons per $(\lambda^2/2\pi)$. This metric gives the number of photons needed to actuate a switch with a transverse dimension that has been reduced as much as possible—the diffraction limit of the interacting beams (5, 11). For the spot size of the switching beam used in our experiment ($1/e$ intensity radius of 166 μm), we find that $E \sim 4.4 \times 10^{-2}$ photons/ $(\lambda^2/2\pi)$, corresponding to 11 zeptoJ/ $(\lambda^2/2\pi)$.

Similar behavior is observed for the lower switching power, as shown in Fig. 3C. In this case, a weak switching beam controls output beams that are 6500 times stronger. Even though there is only partial switching, the contrast-to-noise ratio is greater than 10:1 (Fig. 4, D and E, and movie S2). As seen in Fig. 4F, $\tau = 3$ μs , which is somewhat faster than that observed at the higher power, possibly due to the fact that only part of the output light rotates to the new angle. Using this response time, we find $N_p = 2700$ photons, $N_p E_p = 690$ aJ, and $E \sim 3 \times 10^{-3}$ photons/ $(\lambda^2/2\pi)$ [770 yoctoJ/ $(\lambda^2/2\pi)$].

These results demonstrate that a switch based on transverse optical patterns is capable of controlling high-power beams with weak ones, exhibits much higher sensitivity, and can be realized with the use of a simple experimental setup. Such a switch could be used as a binary element in a computation or communication system because, at high

switching beam power, our device operates like a transistor used in digital logic (i.e., the transistor's output is either off or saturated). Additionally, this switch could be used as a router if information is impressed on the output light (e.g., by modulating the pump beams).

For comparison, the sensitivity of our switch, characterized by E , far exceeds that demonstrated by other methods such as electromagnetically induced transparency (EIT). To date, the best EIT-based switch results have been reported by Braje *et al.* (8) who have achieved $E \sim 23$ photons/ $(\lambda^2/2\pi)$; our switch is more than 5000 times as sensitive. Furthermore, the EIT experimental setup is much more complicated, requiring the use of cooled and trapped rubidium atoms and limited to output beams that are much weaker than the input switching beam. Recently, a transient switch based on laser beams propagating through a warm rubidium vapor in a simple setup has been reported, but it does not operate in the low-light-level regime (10).

Although speculative, our technique might be useful at telecommunication wavelengths where high-quantum efficiency, low-noise, single-photon detectors are difficult to realize. For these wavelengths, the rubidium vapor could be replaced with a molecular gas, such as acetylene or hydrogen cyanide, both of which have resonances in the telecommunication band. A low-light-level incident beam could then be detected by switching the output beam onto a standard telecommunications-band detector.

In addition, our general method of exploiting the sensitivity of patterns to small perturbations may find application in other scientific disciplines. For example, modulational instabilities, which often give rise to pattern formation, have been observed in matter waves created with ultracold quantum gases (30), suggesting that atom switching might

be possible by perturbing the gas with a few injected atoms.

Our results may also have implications concerning the fundamental limits of general-purpose computation devices. Many years ago, Keyes (11) realized that thermal energy dissipation places limits on the operational speed of a logic element. By assuming that a saturation-based optical switch has $E \sim 1$ photon/ $(\lambda^2/2\pi)$, he concluded that optical logic elements are limited to rates below 10^{10} s^{-1} . Our observed switching energy density is more than a factor of 300 below that assumed by Keyes, suggesting that optical devices might surpass his estimated limit.

References and Notes

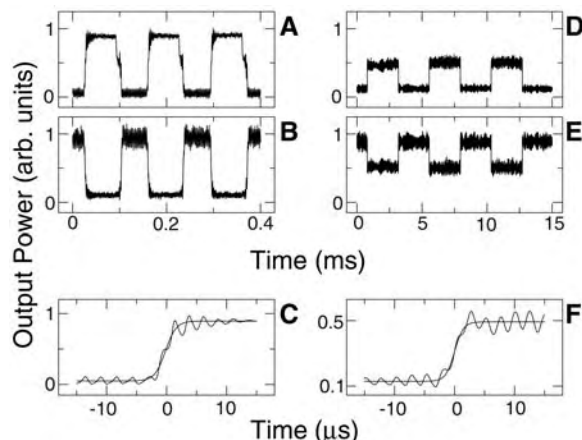
1. R. Ramaswami, K. N. Sivarajan, in *Optical Networks: A Practical Perspective* (Morgan Kaufmann, San Francisco, CA, ed. 2, 2002), ch. 12.
2. H. M. Gibbs, *Optical Bistability: Controlling Light with Light* (Academic Press, Orlando, FL, 1985).
3. M. D. Lukin, *Rev. Mod. Phys.* **75**, 457 (2003).
4. H. Schmidt, A. Imamoglu, *Opt. Lett.* **21**, 1936 (1996).
5. S. E. Harris, Y. Yamamoto, *Phys. Rev. Lett.* **81**, 3611 (1998).
6. M. D. Lukin, A. Imamoglu, *Phys. Rev. Lett.* **84**, 1419 (2000).
7. M. Yan, E. G. Richey, Y. Zhu, *Phys. Rev. A* **64**, 041801 (2001).
8. D. A. Braje, V. Balić, G. Y. Yin, S. E. Harris, *Phys. Rev. A* **68**, 041801 (2003).
9. K. J. Resch, J. S. Lundeen, A. M. Steinberg, *Phys. Rev. Lett.* **89**, 037904 (2002).
10. I. Novikova, A. S. Zibrov, D. F. Phillips, A. André, R. L. Walsworth, *Phys. Rev. A* **69**, 061802 (2004).
11. R. W. Keyes, *Science* **168**, 796 (1970).
12. C. P. Smith *et al.*, *Opt. Commun.* **102**, 505 (1993).
13. R. Martin, A. J. Scroggie, G.-L. Oppo, W. J. Firth, *Phys. Rev. Lett.* **77**, 4007 (1996).
14. F. Prati, M. Travagnin, L. A. Lugiato, *Phys. Rev. A* **55**, 690 (1997).
15. S. Barland *et al.*, *Nature* **419**, 699 (2002).
16. B. Gütlich, R. Neubecker, M. Kreuzer, T. Tschudi, *Chaos* **13**, 239 (2003).
17. Materials and methods are available as supporting material on Science Online.
18. L. A. Lugiato, *Chaos Solitons Fractals* **4**, 1251 (1994).
19. S. E. Harris, *Appl. Phys. Lett.* **9**, 114 (1966).
20. A. Yariv, D. M. Pepper, *Opt. Lett.* **1**, 16 (1977).
21. Y. Silberberg, I. Bar-Joseph, *J. Opt. Soc. Am. B* **1**, 662 (1984).
22. G. Khitrova, J. F. Valley, H. M. Gibbs, *Phys. Rev. Lett.* **60**, 1126 (1988).
23. A. Maître, A. Petrossian, A. Blouin, M. Pinard, G. Grynberg, *Opt. Commun.* **116**, 153 (1995).
24. W. J. Firth, A. Fitzgerald, C. Paré, *J. Opt. Soc. Am. B* **7**, 1087 (1990).
25. D. J. Gauthier, M. S. Malcuit, A. L. Gaeta, R. W. Boyd, *Phys. Rev. Lett.* **64**, 1721 (1990).
26. S. Zibrov, M. D. Lukin, M. O. Scully, *Phys. Rev. Lett.* **83**, 4049 (1999).
27. C. H. van der Wal *et al.*, *Science* **301**, 196 (2003).
28. J. McKeever *et al.*, *Science* **303**, 1992 (2004).
29. M. C. Cross, P. C. Hohenberg, *Rev. Mod. Phys.* **65**, 851 (1993).
30. K. E. Strecker, G. B. Partridge, A. G. Truscott, R. G. Hulet, *Nature* **417**, 150 (2002).
31. We thank K. Gibble for discussions of rubidium cells and the DARPA DSO Slow-Light Program, NSF, and the U.S. Army Research Office for financial support.

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5722/672/DC1
Materials and Methods
Figs. S1 and S2
Movies S1 and S2

24 January 2005; accepted 11 March 2005
10.1126/science.1110151

Fig. 4. Dynamical behavior of the low-level all-optical switch. In the plane of the measurement screen, we place one aperture at the center of one of the spots shown in Fig. 3A (the off state of the switch) and another at the center of one of the spots shown in Fig. 3B (the on state). Temporal evolution of the output power passing through (A and C) the on-state aperture for $P_s = 2.5$ nW, (B) the off-state aperture for $P_s = 2.5$ nW, (D and F) the on-state aperture for $P_s = 230$ pW, and (E) the off-state aperture for $P_s = 230$ pW. The fit lines in (C) and (F) are determined using a sigmoidal function. The rise-time of the switch based on this fit is (C) $\tau = 4$ μs for $P_s = 2.5$ nW and (F) $\tau = 3$ μs for $P_s = 230$ pW. arb. units, arbitrary units.



Extracting a Climate Signal from 169 Glacier Records

J. Oerlemans

I constructed a temperature history for different parts of the world from 169 glacier length records. Using a first-order theory of glacier dynamics, I related changes in glacier length to changes in temperature. The derived temperature histories are fully independent of proxy and instrumental data used in earlier reconstructions. Moderate global warming started in the middle of the 19th century. The reconstructed warming in the first half of the 20th century is 0.5 kelvin. This warming was notably coherent over the globe. The warming signals from glaciers at low and high elevations appear to be very similar.

The worldwide retreat of many glaciers during the past few decades is frequently mentioned as a clear and unambiguous sign of global warming (1, 2). Recent glaciometeorological field experiments and modeling studies have led to a much improved understanding of the link between climate processes and glacier mass balance (3, 4). Yet, the climatic information contained in records of glacier geometry, particularly glacier length, has only partly been exploited. This is perhaps due to the nature of the data. Because data points on glacier length are irregularly spaced in time (Fig. 1), the data are more difficult to handle than some other proxies. Therefore, in most temperature reconstructions of the late Holocene climate, glacier records are not included and most information comes from tree rings (5, 6).

Compared with biogenic climate indicators like tree rings, glacier systems react in a relatively simple way to climate change. The transfer function does not change in time and geometric effects can be addressed. Interestingly, many glaciers are found at high elevations. This implies that a climate signal reflected in glacier fluctuations can be studied as a function of height. The recent discussion on the possible discrepancy between surface-temperature observations and satellite measurements (7) and the problems involved in analyzing radiosonde temperature data (8) demonstrates the importance of climate proxies from high-elevation sites.

Although glacier retreat is mentioned in almost all assessments on climate change, the number of systematic studies of longer records is quite small. Some glaciers have been studied in great detail [including Storglaciären, Sweden (9); Nigardsbreen, Norway (10); Rhonegletscher, Switzerland (11); and Untere Grindelwaldgletscher, Switzerland (12)], but the methods used cannot be applied to a large sample because the required input data

are not available. Direct mass-balance observations have been analyzed to estimate the contribution of glaciers to sea-level change (13). Unfortunately, such observations started only in the second half of the 20th century and do not provide information about the transition from the Little Ice Age to the current climatic state.

Here, I present an objective climatic interpretation of glacier length records from all over the world. A linear inverse model provides the basis for an individual treatment of all length records. Differences in the climate sensitivity and response time of glaciers are taken into account.

Records of glacier length were compiled from various sources, building on a data set from an earlier study (14). It was possible to extend the set of 48 records to a set of 169 records from glaciers found at widely differing latitudes and elevations. The core of the data set comes from the files of the World Glacier Monitoring Service in Zürich (15). Records were then included from glaciers in Patagonia (16), southern Greenland (17), Iceland (18), and Jan Mayen (19). Additional

information was taken from the Satellite Image Atlas of Glaciers of the World (20) and from reports of the Swiss Academy of Sciences (21). The character of the records differs widely (Fig. 1). Some start in 1600 and have typically 10 data points until 1900 and more afterward. Other records start around 1900 but have annual resolution throughout. The longest record is that of the Untere Grindelwaldgletscher, which starts in 1534 (22).

Data points in the earlier parts of glacier records are sparse but normally quite reliable. The information on maximum stands from sketches, etches, paintings, and photographs can be checked with moraine systems that are still in place today. However, records based on information from moraines dated by lichenometry or fossil wood without any additional evidence have not been used in the present study.

The records are not spread equally over the globe. There is a strong bias toward the European Alps, where a wealth of documents exists and glacier monitoring was introduced relatively early. Fluctuations of some glaciers in Iceland and Scandinavia before 1800 have also been documented well (18, 23, 24). Glacier records in North America have not been kept up to date and many series do not extend beyond 1985. The 169 glaciers in the data set are located in the European Alps (93 records), Caucasus (8), tropical Africa (5), Central Asia (9), Irian Jaya (2), New Zealand (2), Patagonia (6), Northwest America (27), South Greenland (1), Iceland (4), Jan Mayen (1), Svalbard (3), and Scandinavia (8). In discussing the results, glaciers are grouped into regions, which are referred to as the Southern Hemisphere (tropics, New Zealand, and Patagonia), Northwest America (mainly Canadian Rockies), the Atlantic sector (South Greenland, Iceland, Jan Mayen, Svalbard, and Scan-

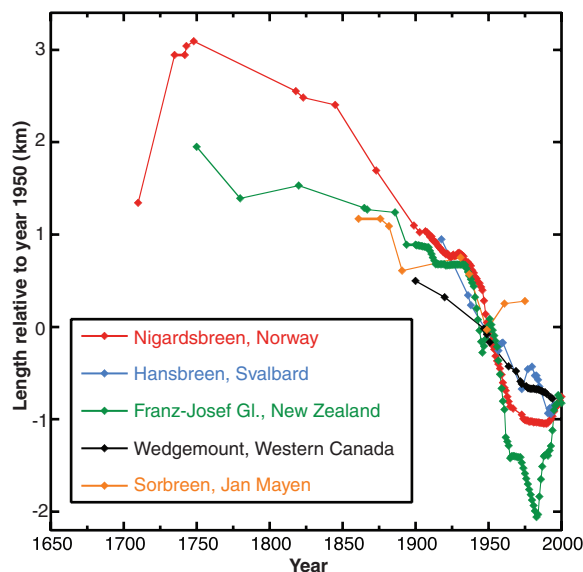
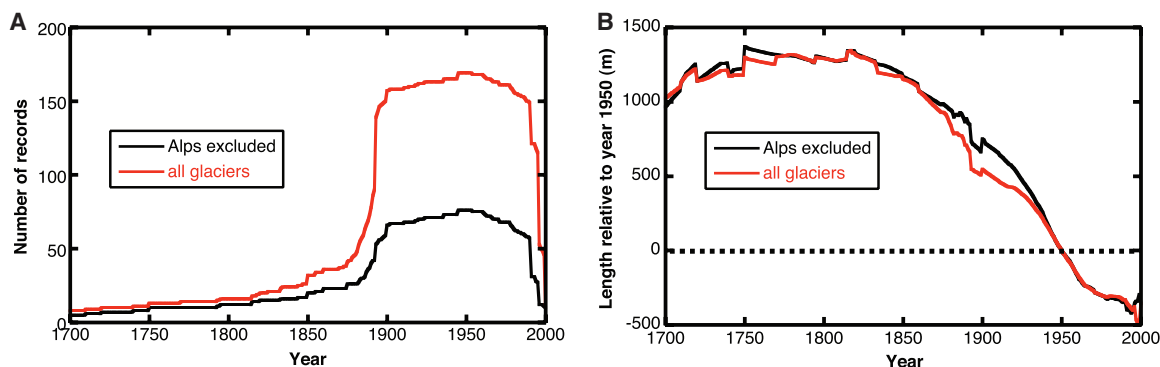


Fig. 1. Examples of glacier length records from different parts of the world. Each dot represents a data point. Data points are scarce before 1900; after 1900 a considerable number of records have annual resolution.

Fig. 2. (A) Number of records for the last 300 years. The decline after 1990 is due to a large delay in the reporting and publishing of data in a suitable form. **(B)** Stacked records of glacier length. Irregularities occur when a glacier with a large length change is added. However, this does not necessarily involve a large change in climatic conditions because glaciers exhibiting large changes are normally those that have a large climate sensitivity (and thus respond in a more pronounced way to, for instance, a temperature change). After 1900, the irregularities disappear because the number of glaciers in the sample increases strongly.



dinavia), and the Alps and Asia (Caucasus and Central Asia).

In all records, data points were connected by interpolation. Normally cubic splines were used. However, in some cases a combination of linear interpolation and fitting with splines performed better. All interpolated records were visually checked for spurious effects.

The number of records reveals a strong increase at the end of the 19th century, both for the Alps and for all other glaciers (Fig. 2A). Stacking all records yields a curve for the change in mean glacier length (Fig. 2B). The curve was not smoothed, implying that some irregularities occur when a glacier with a large change in length is added to the sample (or disappears from the sample at the very end). The curve for all glaciers outside the Alps is notably similar to the curve for the entire sample. This reflects the notion that glacier retreat on the century time scale is rather uniform over the globe. Around 1800, mean glacier length was decreasing and this decrease accelerated gradually. The present data set thus suggests that the Little Ice Age was at its maximum around 1800 rather than at the end of the 19th century as indicated by some other temperature proxies (6).

Histograms of mean retreat rates for selected periods show that very few glaciers in the sample actually became longer (fig. S1). For the period from 1860 to 1900 (36 records), one glacier advanced and all of the others retreated. For the period from 1900 to 1980, 142 of the 144 glaciers retreated.

The response of a glacier to climate change depends on its geometry and on the climatic setting. To unravel the climate signal contained in the glacier length records, it is necessary to discriminate with respect to the climate sensitivity c and to the response time τ (the time a glacier needs to approach a new equilibrium state). Similar to other climate proxies, glacier length fluctuations are the product of variations in more than one meteorological parameter. Glacier mass balance depends mainly on air temperature, solar radiation, and precipitation. Extensive meteorological

experiments on glaciers have shown that the primary source for melt energy is solar radiation but that fluctuations in the mass balance through the years are mainly due to temperature and precipitation (25, 26). Mass-balance modeling for a large number of glaciers has shown that a 25% increase in annual precipitation is typically needed to compensate for the mass loss due to a uniform 1 K warming (3, 27). These results, combined with evidence that precipitation anomalies normally have smaller spatial and temporal scales than those of temperature anomalies (2), indicate that glacier fluctuations over decades to centuries on a continental scale are primarily driven by temperature. Here, the climate sensitivity c is therefore defined as the decrease in equilibrium glacier length per degree temperature increase.

The simplest approach that deals with lag effects is a linear response equation:

$$\frac{dL'(t)}{dt} = -\frac{1}{\tau} [cT'(t) + L'(t)] \quad (1)$$

Here, t is time, L' is the glacier length with respect to a reference state, and T' is a temperature perturbation (annual mean) with respect to a reference state. The inverse model is now obtained by solving for T' :

$$T'(t) = -\frac{1}{c} \left[L'(t) + \tau \frac{dL'(t)}{dt} \right] \quad (2)$$

For any glacier length record, the corresponding temperature history can be obtained with Eq. 2 once the climate sensitivity and response time are known.

A number of glaciers have been studied by explicit numerical modeling (28) or more refined linear inverse modeling (29). However, the input data required for these methods is not available for most glaciers considered here. Therefore, c and τ were determined from a simple theory of glacier dynamics, calibrated with results from numerical studies (30). Climate sensitivity depends in particular on the surface slope (a geometric effect) and the annual precipitation (a mass-balance effect).

Glaciers in a wetter climate are more sensitive (31, 3), and this is taken into account (30). As a result, in the sample of 169 glaciers, c varies by a factor of 10, from ~ 1 to $\sim 10 \text{ km K}^{-1}$ (fig. S2). The response time is to a large extent determined by the slope and the balance gradient (the rate at which the mass gain or loss changes with elevation). Values of τ vary from about 10 years for the steepest glaciers to a few hundreds of years for the largest glaciers in the sample with a small slope (the glaciers in Svalbard). Most of the values are in the range of 40 to 100 years (fig. S2).

Because the right-hand side of Eq. 2 contains the time derivative of glacier length, the calculated temperature curves can be noisy. This noise was removed by filtering, which effectively smoothed out the variability on time scales shorter than about a decade.

Reconstructed temperatures for five regions are shown in Fig. 3A. The interpretation of the temperature curves before 1800 should be done with caution, because the number of records is small (Fig. 2A). From 1860 onward, most regions show a temperature increase. In the first half of the 20th century the temperature rise is notably similar for all regions: about 0.5 K in 40 years. After 1945, the global mean temperature drops slightly until 1970, when it starts to rise again. For North America, the reconstruction shows a marked cooling after 1940, which seems to be at odds with the substantial retreat observed for most glaciers during the past 20 years. This apparent discrepancy is due to the fact that many records from North America are not up to date and end in the period of 1975 to 1990 (table S1).

The global mean temperature shown in Fig. 3A is a weighted mean for the period from 1834 to 1990. To obtain a curve for the period from 1600 to 1990, it can be combined with a stacked temperature reconstruction for all glaciers before 1834 (Fig. 3B). The major sources of error for the global mean temperature reconstruction are (i) the influence of meteorological variables other than temperature, (ii) uncertainty in the climate sensitivity of the glaciers (which affects the amplitude

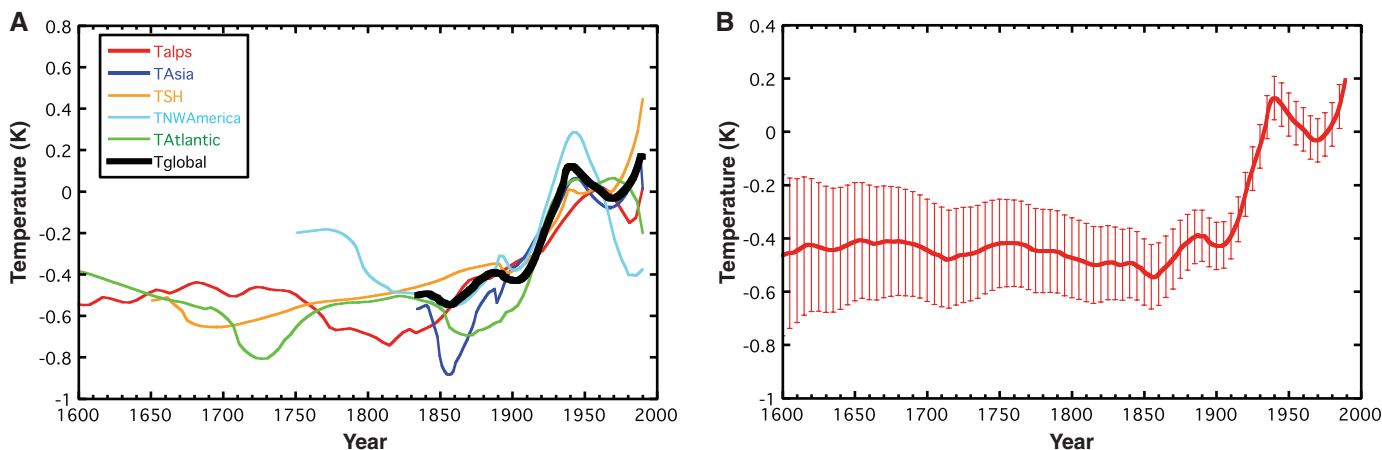


Fig. 3. (A) Temperature reconstruction for various regions. The black curve shows an estimated global mean value, obtained by giving weights of 0.5 to the Southern Hemisphere (SH), 0.1 to Northwest America, 0.15 to the Atlantic sector, 0.1 to the Alps, and 0.15 to Asia.

(B) Best estimate of the global mean temperature obtained by combining the weighted global mean temperature from 1834 with the stacked temperature record before 1834. The band indicates the estimated standard deviation.

of the temperature reconstruction), (iii) uncertainty in the response times (which affects the phase of the temperature reconstruction), and (iv) the number of glacier length records and degree of global coverage. An error estimate was made by assigning $\pm 20\%$ errors to climate sensitivities and response times of individual glaciers and by making 10 subsamples of glacier length records (randomly removing 50% of the records). In total, 100 alternative temperature reconstructions were generated. The standard deviation calculated from this set of reconstructions is taken as an estimate of the error in the best estimate based on all records. This standard deviation has been smoothed in time. Changes in the resulting bandwidth reflect first of all the effect of the steadily increasing number of glacier records (Fig. 3B). The possibility that changes in precipitation are responsible for part of the observed glacier fluctuations cannot be excluded. However, a very large drying on a global scale would be needed to explain the worldwide glacier retreat, and there is no independent evidence at all of such a phenomenon (2).

The derived global temperature record is in broad agreement with other reconstructions and for the last part also with the instrumental record (fig. S3). However, the glacier reconstruction shows a somewhat larger amplitude on the century time scale. Because glaciers need time to react and the number of records drops sharply after 1995, the warming seen in the instrumental record over the past 15 years is not yet reflected in the reconstruction.

Temperature curves appear to be very similar for glaciers with low and high median elevation (fig. S4A). Low and high glaciers are classified as glaciers with median elevation below and above 2850 m, respectively (fig. S5). For this threshold, the number of

glaciers in both classes is approximately equal. Although the evidence is not conclusive because only a limited altitudinal range is considered, the glacier record does not show any sign of a height dependence of the global warming signal.

Glaciers witnessed a particularly strong warming at high northern latitudes in the first decades of the 20th century (fig. S4B). Unlike the global mean signal, reconstructed temperature shows a minimum in the second half of the 19th century, when the Northern Hemisphere mid-latitudes were already warming up.

The temperature reconstruction presented here is fully independent of other sources (proxy or instrumental). It thus provides complementary evidence on the magnitude of the current global warming, on the time that this warming started, and on the notion that in the lower troposphere the warming appears to be independent of elevation.

References and Notes

1. M. B. Dyurgerov, M. F. Meier, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1406 (2000).
2. C. K. Folland, T. R. Karl, in *Climate Change 2000: The Scientific Basis*, J. T. Houghton et al., Eds. (Cambridge Univ. Press, Cambridge, 2001), pp. 101–181.
3. J. Oerlemans, *Glaciers and Climate Change* (A. A. Balkema Publishers, Rotterdam, Netherlands, 2001).
4. C. Vincent et al., *J. Geophys. Res.* **109**, 10.1029/2003JD003857 (2004).
5. K. R. Briffa et al., *Holocene* **12**, 737 (2002).
6. M. E. Mann, P. D. Jones, *Geophys. Res. Lett.* **30**, 10.1029/2003GL017814 (2003).
7. J. Christy, R. Spencer and W. Braswell, *J. Atmos. Oceanic Technol.* **17**, 1153 (2000).
8. J. R. Lanzante, S. A. Klein, *J. Clim.* **16**, 224 (2003).
9. A. P. Stroeven, *Geogr. Ann. Ser. A Phys. Geogr.* **78**, 133 (1996).
10. J. Oerlemans, *Ann. Glaciol.* **24**, 382 (1997).
11. J. Wallinga and R. S. W. van de Wal, *J. Glaciol.* **44**, 383 (1998).
12. M. J. Schmeits, J. Oerlemans, *J. Glaciol.* **43**, 152 (1997).
13. M. B. Dyurgerov, M. F. Meier, *Arctic Alp. Res.* **29**, 379 (1997).
14. J. Oerlemans, *Science* **264**, 243 (1994).
15. W. Haeberli, in *Into the Second Century of World*

Glacier Monitoring—Prospects and Strategies, W. Haeberli, M. Hoelzle, S. Suter, Eds. (UNESCO Publishing, Paris, 1998), pp. 35–51.

16. M. Aniya, *Bull. Glaciol. Res.* **18**, 55 (2001).
17. A. Weidick, in *Gletschere i Sydgrønland* (Greenland Geological Survey, Copenhagen, 1988), pp. 47–49.
18. O. Sigurdsson, *Jökull* **45**, 3 (1998).
19. E. Anda, O. Orheim, J. Mangerud, *Polar Res.* **3**, 129 (1985).
20. R. J. Williams, J. G. Ferrigno, Eds., *Satellite Image Atlas of Glaciers of the World* (U.S. Geological Survey, Washington, DC, 1989–1995), vols. 1–7.
21. Glaciological Commission of the Swiss Academy of Sciences, *Glaciological Reports 1–122* (Glaciological Commission of the Swiss Academy of Sciences, Bern, 1893–2001).
22. H. J. Zumbühl, *Die Schwankungen der Grindelwaldgletscher in den historischen Bild- und Schriftquellen des 12. bis 19. Jahrhunderts* (Birkhäuser Verlag, Basel, 1980).
23. G. Østrem, O. Liestøl, B. Wold, *Nor. Geogr. Tidsskr.* **30**, 187 (1976).
24. J. Bogen, B. Wold, G. Østrem, in *Glacier Fluctuations and Climatic Change*, J. Oerlemans, Ed. (Kluwer, Dordrecht, Netherlands, 1989), pp. 305–323.
25. W. Greuell, C. J. J. P. Smeets, *J. Geophys. Res.* **106**, 31,717 (2001).
26. J. Oerlemans et al., *Bound.-Layer Meteor.* **92**, 3 (1999).
27. R. J. Braithwaite, Y. Zhang, *J. Glaciol.* **46**, 7 (2000).
28. J. Oerlemans et al., *Clim. Dyn.* **14**, 267 (1998).
29. E. J. Klok, J. Oerlemans, *Holocene* **13**, 343 (2003).
30. Climate sensitivity c and response time τ are modeled as $c = 2.3P^{0.6s^{-1}}$ and $\tau = 13.6\beta^{-1s^{-1}}(1 + 20s)^{-1/2}L^{-1/2}$. Here, P is the climatological annual precipitation in m/year, s is the mean surface slope of the glacier, β is the altitudinal mass-balance gradient, and L is the glacier length in the reference case in meters. These formulations were obtained by calibrating a simple model of glacier dynamics ([3], pages 61 and 73) with more explicit numerical modeling of a limited set of glaciers ([3], pages 83 to 92). Values for P were obtained from climate stations and climate atlases. The balance gradient has been related to P by $\beta = 0.006P^{1/2}$ (here P is in m/year).
31. M. F. Meier, *Science* **226**, 1418 (1984).
32. I thank all colleagues that helped to collect data, in particular E. J. Klok.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1107046/DC1
 Figs. S1 to S5
 Tables S1 and S2
 References

2 November 2004; accepted 18 January 2005
 Published online 3 March 2005;
 10.1126/science.1107046
 Include this information when citing this paper.

Early Local Last Glacial Maximum in the Tropical Andes

Jacqueline A. Smith,^{1*} Geoffrey O. Seltzer,^{1†} Daniel L. Farber,²
Donald T. Rodbell,³ Robert C. Finkel⁴

The local last glacial maximum in the tropical Andes was earlier and less extensive than previously thought, based on 106 cosmogenic ages (from beryllium-10 dating) from moraines in Peru and Bolivia. Glaciers reached their greatest extent in the last glacial cycle ~34,000 years before the present and were retreating by ~21,000 years before the present, implying that tropical controls on ice volumes were asynchronous with those in the Northern Hemisphere. Our estimates of snowline depression reflect about half the temperature change indicated by previous widely cited figures, which helps resolve the discrepancy between estimates of terrestrial and marine temperature depression during the last glacial cycle.

It is unclear whether the last glacial maximum (LGM) was a globally synchronous event or whether a local last glacial maximum (LLGM) occurred in the tropics at a different time (1). This issue has broad implications for our understanding of the forcings that underlie the massive global climate changes associated with glacial cycles, and it underpins all paleoclimate reconstructions for the tropics. An absolute chronology for the last glacial cycle and ice coverage in the tropics is unavailable, largely because of the limitations of radiocarbon dating and the scarcity of datable material in glacial environments above 4000 m above sea level (masl). Existing glacial chronologies for the tropics are based primarily on minimum-limiting radiocarbon dates and relative dating techniques (2). An absolute chronology would have important implications for paleoclimate reconstructions and the debates over the synchronicity of glaciation in the Northern and Southern Hemispheres (3), the synchronicity of continental and mountain glaciation (4), the magnitude of tropical climate change at the LGM (5–7), and the relation between sea surface temperatures (SSTs) and temperatures over continental regions (8). Snowline depression of ~1000 m at a presumed LGM age of ~21 thousand years before the present (ky B.P.) is often cited as evidence for terrestrial temperatures 5° to 6°C cooler than at present in the tropics (6). However, the climatic interpretation of snowline depression is complicated (9), because the timing of

glacial coverage in the tropics has been poorly constrained. Here we present an absolute chronology of the LLGM based on cosmogenic dating (¹⁰Be) of erratics on moraines in the tropical Andes of Peru and Bolivia.

We selected field areas that allowed direct comparison of our terrestrial glacial chronology to existing lacustrine paleoclimate records from nearby lakes: Lake Junin (11°S, 76°W) in central Peru and Lake Titicaca (16°S, 69°W) on the Peruvian-Bolivian Altiplano (Fig. 1). Both lakes predate the last glacial-interglacial transition and have provided valuable paleoclimate proxy records in the form of sediment cores (10–14). Lake Junin and Lake Titicaca lie on high plateaus (>3800 masl) between the eastern and western cordillera of the Andes and receive moisture predominantly from easterly winds crossing the Amazon Basin during the austral summer (15). We dated moraines in four valleys in the eastern cordillera bordering Lake Junin and in Milluni and Zongo Valleys in the Cordillera Real, approximately 35 km east of Lake Titicaca (fig. S1 and supporting online text). We used the concentration of cosmogenic ¹⁰Be to deter-

mine the exposure ages of boulders on moraines in the study areas (Table 1 and table S1) (16, 17). Analytical uncertainties on the entire set of exposure ages average $3.0 \pm 1.5\%$ (17). We assigned dated moraines to one of four moraine groups (A to D) based on an assessment of the position of the moraine within the valley, the geomorphic character of the moraine, and the exposure ages obtained from the moraine.

In the Junin valleys (Fig. 2A and fig. S1, A to D), 146 ¹⁰Be exposure ages clearly distinguish moraines deposited late in the last glacial cycle from moraines that are considerably older (Fig. 3). Exposure ages of boulders on LLGM moraines are typically 34 to 22 ky B.P.; ages on older moraines generally predate the last glacial cycle (>65 ky B.P.) and in many cases are dated >200 ky B.P. (18). In the west-facing valleys, the LLGM moraine is the lower of two end moraines that are located midway downvalley; a younger (~20 to 16 ky B.P.) moraine typically dams a lake (or paleolake) 1 to 2 km upvalley from the LLGM ice limit.

In the Cordillera Real (Fig. 2B and fig. S1, E and F), all 42 of the ¹⁰Be exposure ages fall within the last glacial cycle (table S1). In Milluni Valley, exposure ages on two sharp-crested parallel lateral moraines sampled at ~4500 to 4600 masl lie between 34 and 23 ky B.P. (with two younger outliers). Exposure ages on a recessional moraine at ~4650 masl are late glacial to early Holocene (18 to 9 ky B.P.). All exposure ages from Zongo Valley, even those from the lateral moraines at 3400 to 3500 masl, are late glacial (<20 ka) or younger (19).

The pattern of LLGM moraine ages in the Junin region is consistent from valley to valley, whereas the pattern in the Cordillera Real is not. Exposure ages in the Junin region suggest that glaciers reached their maximum extent of the last glacial cycle as early as ~34 ky B.P. Recession was under way by ~21 ky B.P., followed by stillstand or readvance

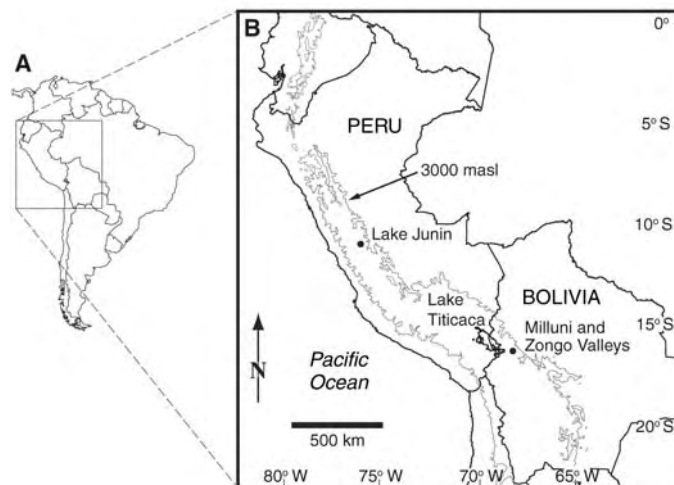


Fig. 1. Site location maps of the study areas. (A) South America, with area covered by (B) indicated. (B) Map of Peru, Bolivia, and adjacent countries, showing the general location of Lake Junin and the Junin valleys, Lake Titicaca, and Milluni and Zongo Valleys in the Cordillera Real. The 3000-m contour line bounds the eastern and western cordillera of the Andes and the intervening high plateaus, including the Altiplano.

¹204 Heroy Geology Laboratory, Department of Earth Sciences, Syracuse University, Syracuse, NY 13244–1070, USA. ²Lawrence Livermore National Laboratory, Post Office Box 808, L-201, Livermore, CA 94551, USA. ³Department of Geology, Olin Building, Union College, Schenectady, NY 12308, USA. ⁴Center for Accelerator Mass Spectrometry, MS L-397, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550–9234, USA.

*To whom correspondence should be addressed. E-mail: jasmit10@syr.edu
†Deceased.

between ~16 and 20 ky B.P. In Milluni Valley, the maximum extent of ice during the last glacial cycle occurred ~34 to 23 ky B.P., with subsequent recession and several stillstands or readvances (20). In Zongo Valley, however, late-glacial ice extended downvalley more than 1 km lower in altitude than ice of similar age in Milluni Valley. We have not yet identified a moraine older than 20 ky B.P. in Zongo Valley. If the LLGM terminus descended past 3400 masl to ~3000 masl, as seems likely from geomorphic evidence, Müller's (21) estimate of ~4100 to 4200 masl for the LLGM limit was too high by at least 700 m and perhaps by more than 1 km.

Our glacial chronology is consistent with the paleoclimate proxy records from Lake Junin and Lake Titicaca. Seltzer *et al.* (10, 11) used magnetic susceptibility (MS) and stratigraphic records to bracket the LLGM between ~30 ky B.P. and ~21 ky B.P. in a sediment core from Lake Junin with a basal age of ~40 ¹⁴C ky B.P. (radiocarbon years before the present). MS in Lake Junin sediments was elevated (relative to Holocene and late-glacial levels) even before 30 ky B.P., including sharp increases ~37 to 35 and ~33 ky B.P. (10, 11). Seltzer *et al.* (10) identified the glacial-interglacial transition with a dramatic drop in MS ~21 ky B.P. and the onset of full deglaciation with the increase in organic

carbon ~16 ky B.P. Our record from the Junin valleys suggests maximum ice extent ~34 to 22 ky B.P., ice retreat from the terminal moraine by 21 ky B.P., formation of a second end moraine upvalley between ~20 and 16 ky B.P., and steady retreat after ~15 ky B.P. Lake Titicaca sediment cores indicate a wet glacial period and the onset of deglaciation ~24 to 19 ky B.P. The LGM and late-glacial period (until ~15 ky B.P.) on the Altiplano were wet (12, 13) and lake levels were high (14). Temperature decreased by ~5°C from 27.5 to 21 ky B.P. and warming began ~21 ky B.P. (14). Seltzer *et al.* (11) identified the glacial-interglacial transition with decreases in MS and sedimentation rates ~24 and ~19.5 ky B.P., respectively, in two sediment cores from Lake Titicaca that show elevated MS before the transition. Our glacial chronology shows maximum ice extent in Milluni Valley ~34 to 23 ky B.P. and construction of a recessional moraine (~10 km upvalley and ~50 m higher in altitude) by 18 ky B.P. Exposure ages suggest that ice remained in Zongo Valley almost into the Holocene.

We estimate that equilibrium-line altitude (ELA) depression at the LLGM was 300 to 600 m in the Junin valleys and Milluni Valley and 800 to 1000 m in Zongo Valley (fig. S2). Even with uncertainty of ±100 m, our estimates of LLGM ELA depression for the Junin valleys and Milluni Valley are about half of published estimates of 900 to 1000 m for the tropical Andes (6, 7). We can use the position of LLGM moraines to constrain paleoclimatic change at the time of moraine deposition. If we assume that the change in ELA (Δ ELA) was entirely a function of a change in temperature (Δ T), we estimate a Δ T of -2° to -4°C in the Junin valleys and Milluni Valley at the LLGM [using a moist adiabatic lapse rate of 6°C/km (7)]. Recent estimates put the LLGM temperature depression in the tropical Andes and Amazon Basin at ~5°C (14, 22) and the change in SST in the eastern equatorial Pacific at -2.8° ± 0.7°C (23). If the Junin and Milluni estimates of Δ ELA accurately reflect the Δ T at the LLGM, as these recent Δ T estimates suggest, then we can infer that the Zongo Valley glacier was anomalously long.

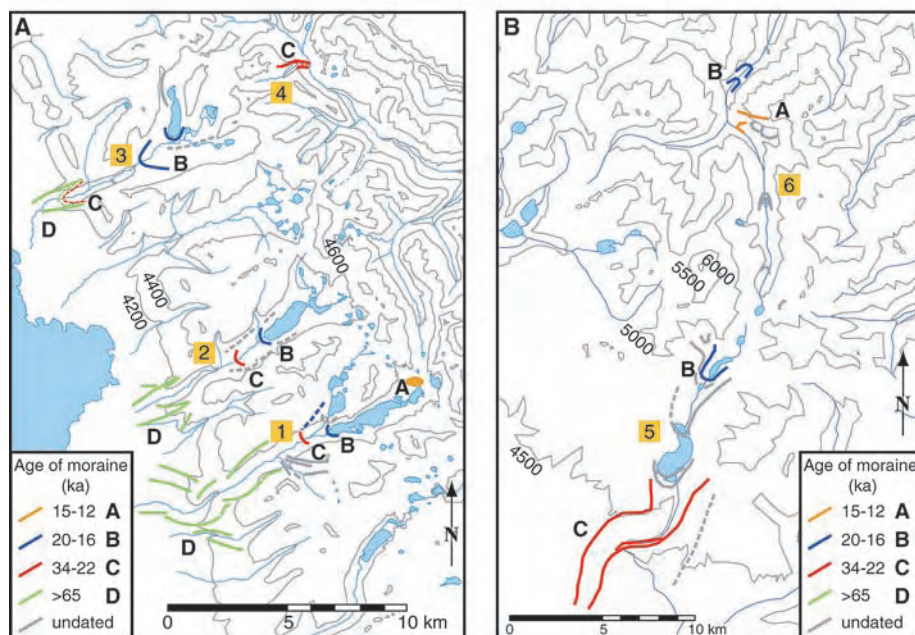


Fig. 2. Exposure ages of moraines in the Junin valleys (A) and Milluni and Zongo Valleys (B) based on cosmogenic dating with ¹⁰Be. Thin gray lines are contour lines, with altitudes in masl; contour intervals are 200 m (A) and 500 m (B). Heavier solid lines indicate distinct moraines observed in the field and/or on aerial photographs; dashed lines indicate less distinct moraines or boulder trains. Ages shown were calculated without erosion or uplift (17). Analytical uncertainties on exposure ages average 3.0 ± 1.5% for all moraine groups combined (moraine groups A, B, C, and D). The LLGM in both areas was reached ~34 ky B.P. (A) The Junin valleys include Alcacocha Valley (1), Antacocha Valley (2), Calcacocha Valley (3), and Collpa Valley (4). In west-facing Junin valleys, LLGM glaciers (moraine group C) were small relative to glaciers that deposited larger moraines downvalley (moraine group D) before the last glacial cycle (>65 ky B.P.). (B) In Milluni Valley (5) and Zongo Valley (6), all moraines that we have dated were deposited during the last glacial cycle (<65 ky B.P.). Contemporaneous moraines lie more than 1000 m lower in Zongo Valley than in Milluni Valley. Detailed sampling locations are illustrated in fig. S1.

Table 1. Exposure ages (¹⁰Be) for moraine groups A to C deposited during the last glacial cycle (17). Statistical data are shown for sample ages excluding outliers (I) and for all sample ages (II). As used here, the term "outliers" refers to exposure ages greater than 2 standard deviations (SD) from the mean age of the moraine group.

Moraine group	(I) Mean age (excluding outliers) (years)	SD of the ages (excluding outliers) (years)	Age range (excluding outliers) (years)	(II) Mean age (all samples) (years)	SD of the ages (all samples) (years)	Age range (all samples) (years)	Number of samples (including outliers)	Number of outliers (>2σ from mean)	Relative SD*	Analytical uncertainty†
A	14,500	2,000	11,500–18,400	15,300	4,300	11,500–31,100	19	1	14%	4.1%
B	16,300	2,900	7,900–22,700	15,900	4,400	3,500–29,700	50	4	18%	4.2%
C	27,800	6,400	15,500–53,200	32,100	27,200	15,500–189,900	37	1	17%	3.2%

*Relative SD = (SD of the ages, excluding outliers)/(mean ¹⁰Be age, excluding outliers). †Analytical uncertainty is the blank-corrected uncertainty in the accelerator mass spectrometer measurements.

The length of the Zongo Valley glacier can be explained by the effects of local factors such as shading by the sheer valley walls (24)

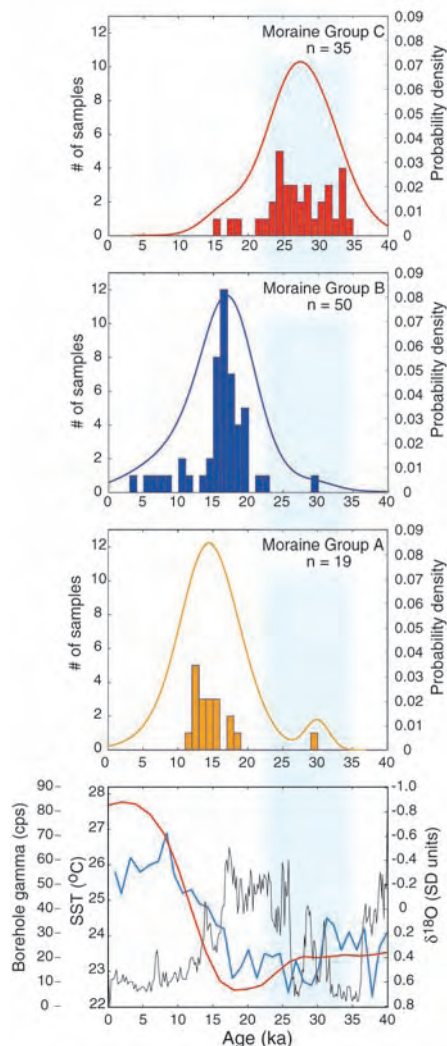


Fig. 3. Exposure ages and paleoclimate proxy records for 0 to 40 ky B.P. The top three panels are histograms and probability density (PD) curves (32) showing age distributions for cosmogenic ¹⁰Be exposure ages in moraine groups C, B, and A, respectively, from the top down; *n* refers to the number of samples included in the histogram and PD curves. Analytical uncertainties on exposure ages average $3.8 \pm 1.4\%$ for moraine groups A, B, and C combined (moraine group D is omitted). Two exposure ages >40 ky B.P. have been omitted from the histogram for moraine group C (see table S1). The bottom panel shows borehole gamma radiation readings [in counts per second (cps)] from the Salar de Uyuni, Bolivia [outer left y axis, black curve (27)], SSTs at the Cocos Ridge reconstructed from Mg/Ca ratios in foraminifer shells from core TR163-19 [inner left y axis, blue curve (23)], and the SPECMAP stacked benthic $\delta^{18}\text{O}$ record [right y axis, red curve (29)]. Higher borehole gamma readings are interpreted as indicating wetter conditions (27), and higher (positive) benthic $\delta^{18}\text{O}$ values are interpreted as indicating greater global ice volume (29). The vertical blue band spanning the four panels marks the LLGM in the tropical Andes.

and debris cover on the ice surface (25), compounded by orographically controlled precipitation resulting from moisture-bearing easterlies (26) and/or the positive feedback set up by ice advancing into a steep, positive precipitation gradient. Thus, our findings sound a note of caution for the use of generalized ELA estimates in paleoclimate reconstructions for areas of complex topography such as the Andes.

We propose that a shifting balance between temperature and precipitation changes promoted glacier development and expansion early in the last glacial cycle and that warming initiated deglaciation ~21 ky B.P. The SST record from the Cocos Ridge in the eastern equatorial Pacific (Fig. 3) shows overall lower SSTs from ~70 to 12 ky B.P., with sharp decreases at ~68 to 65 ky B.P., ~38 ky B.P., ~30 to 25 ky B.P., and ~22 to 16 ky B.P. (23). Published moisture records suggest that the last glacial cycle was wetter than the present (27, 28). A moisture record for the Salar de Uyuni (~20°S) on the southern Bolivian Altiplano (27) indicates a dry period from ~67 to 61 ky B.P., followed by wetter conditions with periods of perennial lakes from ~60 to 15 ky B.P., and finally dry conditions from ~15 ky B.P. to the present. Ice may have begun to accumulate in the highest areas as soon as precipitation increased at ~60 ky B.P. and then expanded after the sharp SST cooling at ~38 ky B.P. (23). With increased precipitation and persistent cool temperatures, glaciers in the highest areas reached their maxima at ~34 ky B.P., while ice began to accumulate and/or expand in the broader Junin valleys with lower headwalls. Glaciers retreated behind their end moraines at ~22 to 21 ky B.P. in response to warming. Glaciers stabilized at their new positions for ~5000 years, from ~20 to 16 ky B.P.; the resulting recessional moraines dammed lakes (or paleolakes) in the west-facing Junin valleys (fig. S1). The scarcity of younger moraines upvalley from these lakes indicates that the subsequent and final deglaciation was steady, leaving the basins relatively free of fluvio-glacial deposits and till. Decreasing precipitation (driven by the precession cycle) and further warming after ~15 ky B.P. made this last deglaciation a rapid one. However, in steep and deep valleys on the eastern side of the eastern cordillera, such as Zongo Valley, deglaciation proceeded differently. There, the combination of sheer valley walls that shaded the glacier; rockfall and erosional debris that insulated the ice; and orographically controlled precipitation that supplied the accumulation area maintained the glacier terminus at an altitude more than 1 km lower than those of contemporaneous termini of paleoglaciers on the western side (fig. S1).

Our results indicate that the LLGM in the tropical Andes began ~34 ky B.P., thus pre-

ceding by as much as ~10 to 12 ky B.P. the maximum continental ice volume of the last glacial cycle as inferred from the marine oxygen isotope record (29). We estimate that ELA depression was typically about half the widely cited value of ~900 to 1000 m, except where local factors enhanced glacier size. The close agreement between our ¹⁰Be-derived chronology and paleoclimate proxy records from nearby lakes shows that these tropical mountain glaciers are sensitive indicators of tropical climate that integrate a combination of climatic variables into a seemingly simple signal. The mountain glaciers of the tropical Andes apparently reached their maximum extent of the last glacial cycle before the continental ice sheets (Fig. 3) (29). Records of mountain glaciations from some other mid- and high-latitude regions have shown similar divergence from global ice volume records (29–31), thus reflecting a complex mosaic of regional climatic patterns during the last glacial cycle.

The asynchronicity in timing of the LLGM in the tropics and ice extent in the Northern Hemisphere implies that the climate-forcing mechanisms were decoupled at some level during the last glacial cycle. This, in turn, challenges conclusions about interhemispheric teleconnections that are based on the assumption that the LGM was globally synchronous (3) and that the magnitude of mountain snowline depression was uniform everywhere.

References and Notes

1. G. O. Seltzer, *Quat. Sci. Rev.* **20**, 1063 (2001).
2. J. A. Smith, G. O. Seltzer, D. T. Rodbell, A. G. Klein, *Quat. Int.*, in press.
3. P. U. Clark, R. B. Alley, D. Pollard, *Science* **286**, 1104 (1999).
4. A. Gillespie, P. Molnar, *Rev. Geophys.* **33**, 311 (1995).
5. CLIMAP, *Geol. Soc. Am. Map Chart Ser. MC-36* (1981).
6. S. C. Porter, *Quat. Sci. Rev.* **20**, 1067 (2001).
7. A. G. Klein, G. O. Seltzer, B. L. Isacks, *Quat. Sci. Rev.* **18**, 63 (1999).
8. D. Rind, D. Peteet, *Quat. Res.* **24**, 1 (1985).
9. G. O. Seltzer, *Quat. Res.* **41**, 154 (1994).
10. G. Seltzer, D. Rodbell, S. Burns, *Geology* **28**, 35 (2000).
11. G. O. Seltzer et al., *Science* **296**, 1685 (2002).
12. P. A. Baker et al., *Science* **291**, 640 (2001).
13. P. M. Tapia, S. C. Fritz, P. A. Baker, G. O. Seltzer, R. B. Dunbar, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **194**, 139 (2003).
14. G. M. Paduano, M. B. Bush, P. A. Baker, S. C. Fritz, G. O. Seltzer, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **194**, 259 (2003).
15. R. Garreaud, M. Vuille, A. C. Clement, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **194**, 5 (2003).
16. J. C. Gosse, F. M. Phillips, *Quat. Sci. Rev.* **20**, 1475 (2001).
17. Details of materials, methods, ¹⁰Be data, and calculations are available as supporting material on Science Online.
18. J. A. Smith, D. L. Farber, R. C. Finkel, G. O. Seltzer, D. T. Rodbell, *Geol. Soc. Am. Annu. Meeting Progr.* **34**, 131 (2002).
19. On the lowest lateral moraine (~3400 masl), the exposure ages of boulders of the same lithology as the adjacent bedrock (granite) are young (~3 to 12 ky B.P.), whereas those of erratic lithologies (schist/quartzite) are late glacial (~15 to 17 ky B.P.), suggesting a localized rockfall component of the boulder population. This was not noted at other sampling locations.

20. The presence of boulders with a span of exposure ages (~34 to 22 ky B.P.) on the LLGM moraines may mean that glacier termini remained close to their maxima for up to ~12,000 years or fluctuated, building composite moraines. It is also possible that younger ages represent the effects of boulder exhumation, movement, or differential boulder erosion.
21. R. Müller, thesis, University of Zürich (1985).
22. M. Stute *et al.*, *Science* **269**, 379 (1995).
23. D. W. Lea, D. K. Pak, H. J. Spero, *Science* **289**, 1719 (2000).
24. J. Oerlemans, in *Glacier Fluctuations and Climatic Change*, J. Oerlemans, Ed. (Kluwer Academic, Dordrecht, Netherlands, 1989), pp. 353–371.
25. D. H. Clark, M. M. Clark, A. R. Gillespie, *Quat. Res.* **41**, 139 (1994).
26. A. M. Johnson, in *World Survey of Climatology*, Vol. 12, W. Schwerdtfeger, Ed. (Elsevier, New York, 1976), pp. 147–218.
27. S. C. Fritz *et al.*, *Quat. Res.* **61**, 95 (2004).
28. P. A. Baker *et al.*, *Nature* **409**, 698 (2001).
29. J. Imbrie *et al.*, in *Milankovitch and Climate, Part I*, A. L. Berger, J. Imbrie, J. Hays, G. Kukla, B. Saltzman, Eds. (Reidel, Dordrecht, Netherlands, 1984), pp. 269–305.
30. L. A. Owen *et al.*, *Quat. Sci. Rev.*, in press.
31. G. D. Thackray, *Quat. Res.* **55**, 257 (2001).
32. T. V. Lowell, *Quat. Sci. Rev.* **14**, 85 (1995).
33. We acknowledge financial and/or material support from the National Geographic Society (grant 7188-02), Lawrence Livermore National Laboratory, the Geological Society of America, Union College, Syracuse University, and NSF (grant ATM-0081517). We are grateful to two

anonymous reviewers whose comments and suggestions substantially improved the manuscript. We dedicate this report to the memory of our coauthor, colleague, and friend, Geoff Seltzer, who was an integral part of the project from beginning to end.

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5722/678/DC1

Materials and Methods

SOM Text

Figs. S1 and S2

Table S1

References

3 November 2004; accepted 25 February 2005

10.1126/science.1107075

Laboratory Earthquakes Along Inhomogeneous Faults: Directionality and Supershear

Kaiwen Xia,^{1,2*} Ares J. Rosakis,^{1†} Hiroo Kanamori,² James R. Rice³

We report on the experimental observation of spontaneously nucleated ruptures occurring on frictionally held bimaterial interfaces with small amounts of wave speed mismatch. Rupture is always found to be asymmetric bilateral. In one direction, rupture always propagates at the generalized Rayleigh wave speed, whereas in the opposite direction it is subshear or it transitions to supershear. The lack of a preferred rupture direction and the conditions leading to supershear are discussed in relation to existing theory and to the earthquake sequence in Parkfield, California, and in North Anatolia.

There is evidence for supershear rupture propagation during earthquakes (1–6), and the link between large earthquakes and the conditions leading to supershear has been established experimentally (7).

Although many of the physical aspects of dynamic rupture (including supershear) are recently becoming clearer in relation to homogeneous faults (i.e., faults separating the same material) (7–13), the behavior of spontaneously nucleated ruptures in inhomogeneous faults (i.e., separating materials with different wave speeds) is still experimentally unexplored except in (14). Because many large earthquakes rupture on faults separating rock masses with different wave speeds, the mechanics of sliding in bimaterial systems is relevant to seismology. Specifically, the questions of whether a preferred rupture direction exists (11), whether a unilateral or bilateral faulting dominates (15), and what the condition is for supershear rupture propagation are particularly relevant to fault dynamics. These questions are

also relevant to hazard potentials of large earthquakes, because directionality and rupture speed control near-field ground motions.

According to analysis and numerics, there are two types of such steady, self-sustained pulses (16, 17). One type corresponds to rupture growth in the direction of slip in the lower wave speed material of the system, and this direction is referred to as the positive direction (10, 11). The rupture pulses belonging to this type are subshear and always propagate with a steady velocity $V = +C_{GR}$, the generalized Rayleigh (GR) wave speed of the system. In this work, we refer to such ruptures as positive GR ruptures and abbreviate them as +GR ruptures. The second type of self-sustained ruptures corresponds to propagation in the negative direction opposite to +GR ruptures (17, 18). Such pulses are supershear and always propagate with a steady velocity that is slightly lower than the P -wave speed of the material with the lower wave speed ($V = -C_p^2$). In the present paper, we will abbreviate such ruptures as $-P_{SLOW}$ ruptures.

Our experiments examined the effect of material contrast on the rupture growth of spontaneously nucleated dynamic ruptures hosted by inhomogeneous, frictional interfaces. These interfaces are held together by static, far-field pressure-shear simulating natural tectonic loads. The experiments mimic natural earthquakes where bimaterial contrast between intact rock masses seldom exceeds

30% in shear wave speeds (10). The experimental setup is similar to that used in our previous study of rupture in homogeneous interfaces (7). This configuration has proven effective in producing accurate, full-field, and real-time information of generic rupture characteristics that can ultimately be related to the rupture behavior of natural fault systems.

A Homalite-100 [Homalite (Division of BIG, Incorporated), Wilmington, DE] plate (material 1, top) and a polycarbonate plate (material 2, bottom) are held together by far-field load, P (Fig. 1). The ratio of shear wave speeds in the materials, $C_S^1/C_S^2 = 1.25$, is chosen to be at the high end of the naturally occurring bimaterial range so that the interfacial phenomena of interest are observable in the experiments. The shear wave speeds (Fig. 1) are directly measured for each material by following the shear wave fronts with the use of high-speed photography and photoelasticity. Photoelasticity, being sensitive to maximum shear stress fields, is perfectly suited for measuring shear wave speeds and for scrutinizing shear-dominated rupture processes in brittle, transparent, and birefringent solids (7). The P -wave speeds were calculated by using measured values of Poisson's ratios ($\nu^1 = 0.35$, $\nu^2 = 0.38$) and by using the directly measured shear wave speeds. Plain strain values were used, because we are interested in processes that take place at short wavelengths near the front of the rupture. An independent measurement of the longitudinal wave (P wave) speeds in the plates using ultrasonic transducers has confirmed these calculated values to within 5%. GR waves for this bimaterial pair exist, and their speed is calculated to be $C_{GR} = 959$ m/s, a value which is close to the shear wave speed of polycarbonate.

The dynamic rupture is triggered by means of an exploding wire mechanism, which simulates a localized pressure release at the desired location of the simulated hypocenter (7). About 32 experiments featuring different angles, α (20° , 22.5° , and 25°), and far-field loading P (10, 13, and 18 MPa) were performed, and the rupture events were repeatedly visualized in intervals of 3 μ s by a digital high-speed camera system used in conjunction with dynamic photoelasticity (7).

¹Graduate Aeronautical Laboratories, California Institute of Technology (Caltech), Pasadena, CA 91125, USA. ²Seismological Laboratory, Caltech, Pasadena, CA 91125, USA. ³Division of Engineering and Applied Sciences and Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138, USA.

*Present address: Division of Engineering, Brown University, Box D, Providence, RI 02912, USA.

†To whom correspondence should be addressed. E-mail: rosakis@aero.caltech.edu

The higher level of angles was limited by the static frictional characteristics of the interface. Depending on P and α , three distinct and repeatable rupture behaviors were observed. In all cases, the two separate, semi-circular traces of the shear waves in the two materials were visible as discontinuities in the maximum shear stress field. The ruptures were always bilateral and became progressively asymmetric with time within the time window of all experiments.

Two distinct rupture tips, one propagating to the west and the other to the east with velocities V^E and V^W , respectively, were identified by a distinct concentration of fringe lines (Fig. 2). For this case (case 1), both tips propagated at subshear velocities, $V^E < V^W < C_S^1 < C_S^2$. Differentiation of the rupture length time histories, obtained from a series of high-speed images, allows for estimation of the rupture velocity histories. The rupture propagating to the west is in the direction of slip of the lower wave speed material (positive direction). Within experimental error, this rupture is found to propagate at a constant velocity close to C_{GR} ($V^W = 950 \text{ m/s} \approx +C_{GR}$). The rupture propagating to the east (negative direction) grows at an almost constant subshear velocity $V^E = -900 \text{ m/s}$, which is slower than the Rayleigh wave speed, C_R^2 , in the slower wave speed material. The observations were similar for smaller α values and compressive P loads as well. In this case, the rupture speed to the east remained sub-Rayleigh ($V^E < C_R^1 < C_R^2$). However, its velocity varied continuously across experiments with different load levels and angles. In particular, smaller α and lower P resulted in V^E being lower fractions of C_R^2 .

A distinct but equally repeatable rupture case (case 2) was observed for higher values of α and P . These conditions correspond to higher values of driving stress or to conditions closer to incipient uniform sliding of the entire interface. A typical example corresponding to $\alpha = 25^\circ$ and $P = 18 \text{ MPa}$ (Fig. 3) shows that the rupture was bilateral with a westward tip trailing behind both shear wave traces. This tip propagated at a constant velocity $V^W \approx +C_{GR}$. This observation is identical to the situation described above in relation to lower values of α and P . The eastward propagating tip, however, is different from the previously described case. This tip propagated with a velocity faster than both the shear wave speeds. Moreover, its structure (Fig. 3) is distinctly different from the structure of westward moving +GR type of rupture. As a conclusive proof of its supershear velocity, two distinct shear shock waves are clearly visible. The magnitude of the velocity of the eastward rupture $|V^E|$ was 1920 m/s , which is $\sim 12\%$ less than the longitudinal wave (P wave) speed, C_P^2 , of the

lower wave speed material. $|V^E|$ is also equal to $1.6C_S^1$ or is slightly higher than $\sqrt{2} C_S^2$.

Both cases described above feature westward propagating ruptures that are of the +GR type. Irrespective of the values of α and P , these ruptures have a constant speed $V^W \approx +C_{GR}$ and they propagate in the positive direction. However, those two cases also feature eastward ruptures that are distinctly different in nature. For sufficiently low P and α , the eastward ruptures, which propagate in the negative direction, are purely subshear within the time window of our experiments. For large enough P and α , however, eastward ruptures propagate with a constant supershear velocity whose magnitude is slightly less than C_P^2 and are thus of the $-P_{SLOW}$ type. To visualize an intermediate situation and a controlled transition from one case to the other within the field of view, we reduced P to 13 MPa (Fig. 4, A and B). Indeed for this case (case 3), Fig. 4 shows a smooth transition from case 1 to case 2 within the same experiment. Although the westward rupture remains of the +GR type throughout the experiment, the eastward rupture jumps from a constant subshear velocity (-910 m/s) to a constant supershear velocity (-1920 m/s) and thus transitions to the $-P_{SLOW}$ type. The rupture length-versus-time plot (Fig. 5) also shows the abrupt transition of the eastward rupture from a subshear velocity to a velocity whose magnitude is slightly less than C_P^2 . This happens at a transition length, L , which is $\sim 25 \text{ mm}$.

The eastward transition behavior of case 3 (Figs. 4 and 5) is similar to the one we have discussed (7) in relation to homogeneous interfaces, whereas the transition length, L , is also a decreasing function of α and P . The ruptures that propagate in the negative direction require a certain minimum rupture length before they become supershear. This observation suggests a link between supershear growth in the negative direction and large earthquakes. In contrast, no such transition was observed for +GR ruptures, irrespective of α , P , and rupture length.

Although it is difficult to determine whether the ruptures are pulse-like, crack-like, or a mixture of the two, the observations confirm the existence of two distinct self-sustained and constant speed rupture modes. These are similar to the ones that have been theoretically and numerically predicted (10, 11, 17, 19–21). In particular, a +GR type of rupture mode is always excited instantaneously in the positive direction. Furthermore, a $-P_{SLOW}$ mode is observed as long as the rupture propagating in the negative direction is allowed to grow to sufficiently long distances from the hypocenter. The triggering of the $-P_{SLOW}$ mode is always preceded by a purely subshear, crack-like rupture whose velocity depends on loading and geometry as well as on the bi-

material characteristics. Therefore, the existence of this preliminary and apparently transient stage is one of the main differences with early numerical and theoretical predictions (11, 16, 17, 19, 20, 22, 23).

Another difference from some of the numerical predictions (11, 17, 20, 21) is the consistent experimental observation of bilateral rupturing. In contrast to the experiments, the above numerical predictions only excite one or the other of the two self-sustained rupture modes (17), giving rise to purely unilateral rupture events. They also favor the triggering of the +GR mode in low wave speed bimaterial systems (11, 20). This kind of preference has led to the labeling of the positive direction as the preferred rupture direction (11). These numerical results are also consistent with the notion of rupture directionality (15), whereas our experiments are not. One exception to this rule is provid-

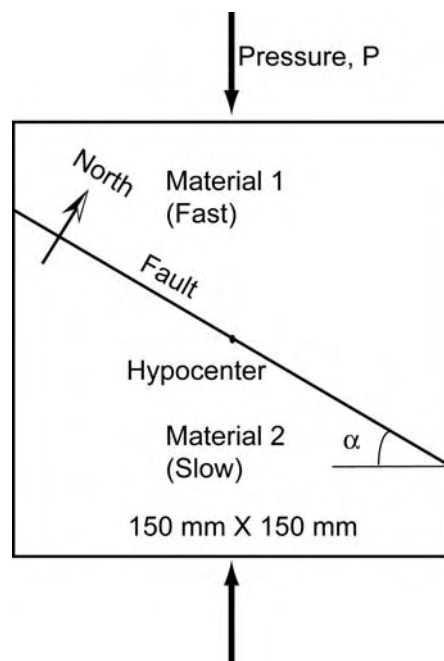


Fig. 1. Laboratory earthquake model composed of two photoelastic plates of the same geometry. The higher wave speed material at the top (Homalite-100) has a density $\rho_1 = 1262 \text{ kg/m}^3$, a shear wave speed $C_S^1 = 1200 \text{ m/s}$, and a longitudinal wave speed $C_P^1 = 2,498 \text{ m/s}$. The lower wave speed material at the bottom (polycarbonate) has a density $\rho_2 = 1192 \text{ kg/m}^3$, a shear wave speed $C_S^2 = 960 \text{ m/s}$, and a longitudinal wave speed $C_P^2 = 2,182 \text{ m/s}$. The fault is simulated by a frictionally held contact interface with an angle to the applied load, which is varied to mimic a wide range of tectonic load conditions. Spontaneous rupture is triggered at the hypocenter through an exploding wire mechanism. The static compressive load P is applied through a hydraulic press. By arbitrary convention, the fault strike runs in the east-west direction with the lower wave speed solid located on the south side. As viewed from the camera, a rupture will produce right lateral slip.

ed by the early numerical analysis by Harris and Day (24), which consistently reports asymmetric bilateral rupture growth in a variety of low speed contrasts in homogeneous fault systems. The results of Harris and Day are qualitatively similar to the experimental observations of cases 1 and 2. However, no transition is reported. To reconcile the observed differences between models and our experiments, we note that unstable slip rupture propagation has also been observed (7) on Homalite/Homalite and polycarbonate/polycarbonate interfaces. Such unstable rupture growth would be possible only if there was a reduction of friction with slip and/or slip rate, and hence such reduction must be a property of both materials when sliding against themselves. It is then plausible to assume that a similar reduction of friction occurs along the Homalite/polycarbonate interface. Hence, its rupture behavior should not be expected to fully correspond to the idealized models of a dissimilar material interface with constant coefficient of friction. Indeed the goal of some of the early theoretical and numerical studies (10, 11, 16, 17, 19, 20, 22, 23) was to investigate what kind of unstable slip would develop on a surface that, as judged from conventional friction notions, was superficially stable in the sense that its friction coefficient, f , did not decrease with slip and/or slip rate. For most brittle solids, however, ample evidence exists that f does decrease with increase of slip and/or slip rate. As a result, a proper model for natural faulting along a bimaterial interface should include both a weakening of f and the slip-normal stress coupling effects of the bimaterial situation. Indeed, such a weakening model

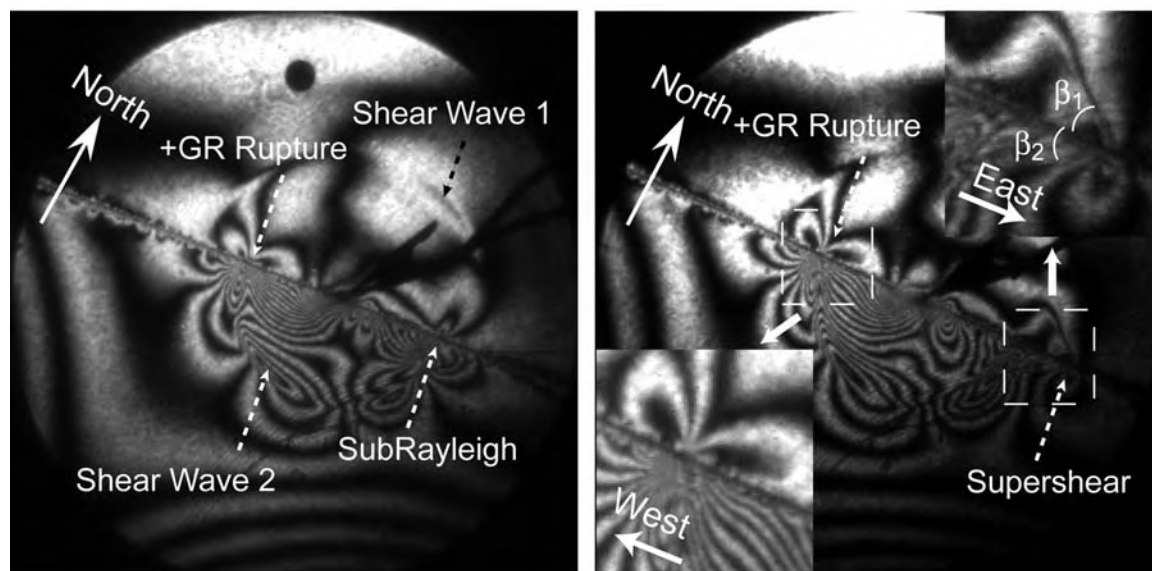
was included by Harris and Day (24). Given the above, it would be an invalid interpretation of the results of the earlier set of papers to conclude that the rupture (including preference for specific rupture mode) scenarios they predict constitute the full set of scenarios available to a real earthquake, of which f decreases with increasing slip and/or slip rate. The consistently bilateral nature of rupture predicted by Harris and Day (24) is perhaps an indication of the effect of including a slip weakening frictional law in their calculations.

Our experiments do not support a preference of rupture direction. Although they support the idea that frictional ruptures in the positive direction always propagate at a specific constant velocity ($V = +C_{GR}$), they still allow for self-sustained intersonic ruptures eventually growing in the negative direction. This possibility becomes more likely if their transient, subshear, and precursory ruptures propagate over a long enough distance and are not arrested prior to transitioning to supershear. The requirement of a critical transition length in the negative direction thus provides a link between large earthquakes and the occurrence of self-sustained supershear rupture in the negative direction.

The 1999, M7.4, Izmit earthquake in Turkey is perhaps a good example for which both modes of self-sustained rupture may have been simultaneously present, as is the case in our experiments. The event was a right-lateral bilateral rupture on a straight segment of the North Anatolian fault. The westward propagating rupture had a speed close to the Rayleigh wave speed, whereas the eastward rupture had a supershear speed that was slightly above the $\sqrt{2}$ times the shear wave

speed of crustal rock (3, 25). Because the geometry of our laboratory earthquake is similar to the Izmit event, direct comparison of the Izmit earthquake and the case described in Fig. 3 reveals similarities. In addition to the right lateral asymmetric bilateral rupture, case 2 featured a subshear westward rupture propagating at $+C_{GR}$. To the east, however, the laboratory rupture propagated at a velocity whose magnitude was slightly lower than C_{GR}^2 , which also happens to be equal to $1.6C_S^1$ for the particular bimaterial contrast of the experiments. Indeed, if one interprets the Izmit event as occurring in an inhomogeneous fault with the lower wave speed material being situated at the southern side of the fault, the field observations and the experimental measurements of both rupture directions and speeds are consistent. Moreover, when the bimaterial contrast is low enough, the differences between C_{GR} and the average of the two Rayleigh wave speeds, $(C_R^1 + C_R^2)/2$, as well as the difference between $1.6C_S^1$ and $\sqrt{2}(C_S^1 + C_S^2)/2$, would be small enough not to be distinguished by the inversion process. In addition, viewing the fault as inhomogeneous can explain the choice of direction for the subshear and the supershear branches, respectively (26). The 1999 Düzce earthquake can also be interpreted along a similar line of argument as used for Izmit. The Düzce rupture featured right lateral slip, and it extended the Izmit rupture zone 40 km eastward through asymmetric bilateral slip (3). Modelling indicates subshear westward and supershear eastward rupture fronts. The direct comparison with case 2 (Fig. 3) thus provides an explanation for the rupture direction and velocity. This explana-

Fig. 2. (Left) Rupture case 1. The photoelastic patterns for an experiment with $\alpha = 22.5^\circ$ and $P = 18$ MPa. Both ruptures to the east and the west are subshear. **Fig. 3. (Right)** Rupture case 2. For $\alpha = 25^\circ$ and $P = 18$ MPa, the bilateral rupture features two distinct tips. The one propagating to the west (positive direction) has a velocity $V^W \approx +C_{GR}$, whereas the one propagating to the east (negative direction) is supershear (V^E). (Upper insert) Two clear lines of discontinuity in the maximum shear contours of photoelasticity. Each of these lines



(shear shock waves) is located at two different angles, $\beta_1 = 41^\circ$ and $\beta_2 = 30^\circ$, to the north and to the south of the fault, respectively. The two angles, β_n , $n = 1, 2$, are related to the shear wave speeds C_S^n and to the rupture velocity V^E , by $\beta_n = \sin^{-1}(V^E/C_S^n)$. This relation

provides independent means of estimating V^E from each individual frame of the high speed camera record without reliance on the less accurate rupture length history. Both methods yield consistent values of $V^E = -1920$ m/s.

Fig. 4. Rupture case 3. Experimental results for $\alpha = 25^\circ$ and $P = 13$ MPa showing transition of the eastward-moving rupture to supershear. The westward rupture retains a constant velocity $V^W \approx +C_{GR}$. (A) Before transition to supershear. (B) After transition to supershear.

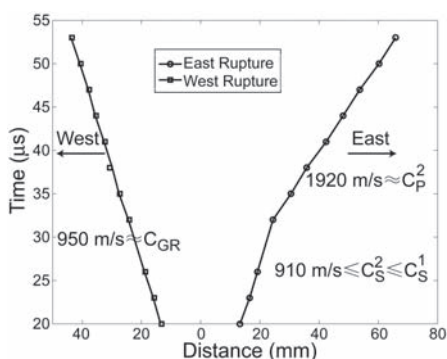
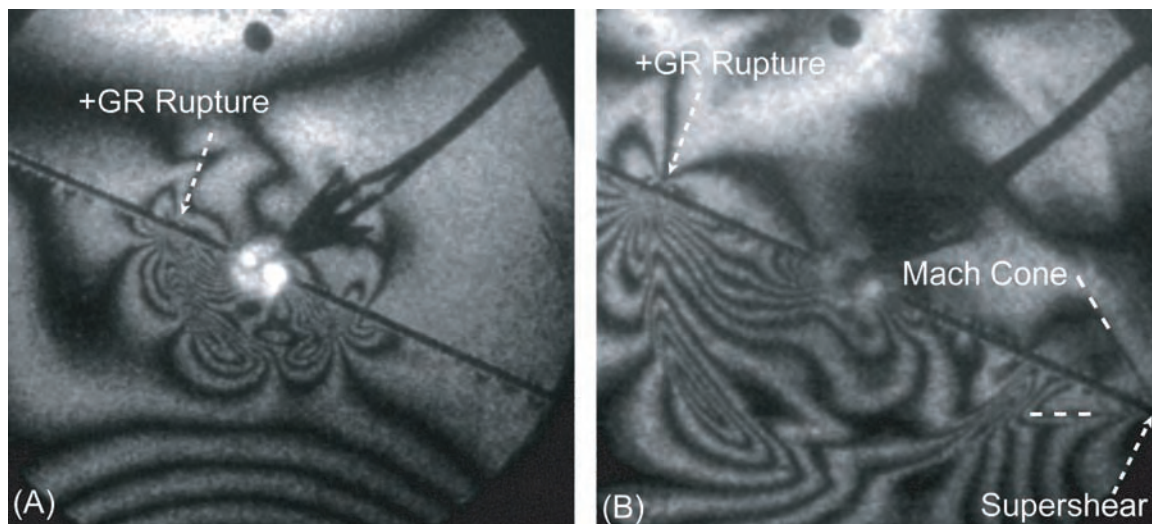


Fig. 5. Rupture time-distance plot for an experiment with $\alpha = 25^\circ$ and $P = 13$ MPa.

tion is plausible if one assumes that the material to the south of the North Anatolian fault, at its western end, is the lower wave speed solid.

By using similar arguments to the ones used for Izmit and Düzce, one can perhaps provide a unified rationalization of the seemingly random rupture directions and rupture velocities of the series of earthquakes that have occurred since 1939 along the North Anatolian fault and ended in 1999 with the Izmit and Düzce events. The following argument requires the assumption that, in average and along its entire length, the North Anatolian fault features the same type of bimaterial inhomogeneity as the one that has been summarized for Izmit and Düzce. Limited evidence supporting such an assumption is currently available (27). If this is true, in some average sense, one would expect that the slight majority (60%) of the large ($M \geq 6.8$) earthquake events (i.e., 1939-M7.9, 1942-M6.9, 1944-M7.5, 1951-M6.8, 1957-M6.8, and 1967-M7.0), which featured westward propagating ruptures, were probably of the +GR type. In other words, that implies that they were classical subshear ruptures that

propagated at $+C_{GR}$ in the positive direction. The remaining four ruptures (i.e., 1943-M7.7, 1949-M7.1, 1999-M7.4, and 1999-M7.1) of the series were irregular in the sense that they featured dominant eastward growth branches, which were probably of the $-P_{SLOW}$ type.

The Parkfield earthquake sequence presents another interesting case in the context of bimaterial rupture. The slip is right-lateral, and the crust on the west side has faster wave speeds than on the east side (28). The two most recent Parkfield earthquakes ruptured the same section of the San Andreas fault. The rupture directions of the 1934 and the 1966 events were southeastward (positive direction), whereas the 2004 earthquake ruptured in the opposite direction (negative direction). Results from early constant friction coefficient studies in bimaterials (16, 17, 19, 20, 22, 23) implies that the positive (southeastward) direction of the 1934 and the 1966 events is a preferred direction. According to this notion, the negative (northwestward) direction of the 2004 earthquake would not be favored. However, our experiments have demonstrated that a rupture in the negative direction can occur.

References and Notes

1. R. J. Archuleta, *J. Geophys. Res.* **89**, 4559 (1984).
2. D. J. Wald, T. H. Heaton, *Bull. Seismol. Soc. Am.* **84**, 668 (1994).
3. M. Bouchon *et al.*, *Geophys. Res. Lett.* **28**, 2723 (2001).
4. M. Bouchon, M. Vallee, *Science* **301**, 824 (2003).
5. W. L. Ellsworth *et al.*, paper presented at the Eleventh International Conference of Soil Dynamics and Earthquake Engineering, Berkeley, California, 7 to 9 January 2004.
6. E. M. Dunham, R. J. Archuleta, *Bull. Seismol. Soc. Am.* **94**, S256 (2004).
7. K. Xia, A. J. Rosakis, H. Kanamori, *Science* **303**, 1859 (2004).
8. A. J. Rosakis, O. Samudrala, D. Coker, *Science* **284**, 1337 (1999).
9. A. J. Rosakis, *Adv. Phys.* **51**, 1189 (2002).
10. J. R. Rice, in *Mechanics for a New Millennium (Proceedings of the 20th International Congress of Theoretical and Applied Mechanics, 27 August to 2*

- September 2000)*, H. Aref, J. W. Phillips, Eds. (Kluwer, Chicago, 2001), pp. 1–23.
11. Y. Ben-Zion, *J. Mech. Phys. Solids* **49**, 2209 (2001).
12. S. M. Day, *Bull. Seismol. Soc. Am.* **72**, 1881 (1982).
13. D. J. Andrews, *J. Geophys. Res.* **81**, 5679 (1976).
14. A. Anooshehpour, J. N. Brune, *Geophys. Res. Lett.* **26**, 2025 (1999).
15. J. J. McGuire, L. Zhao, T. H. Jordan, *Bull. Seismol. Soc. Am.* **92**, 3309 (2002).
16. K. Ranjith, J. R. Rice, *J. Mech. Phys. Solids* **49**, 341 (2001).
17. A. Cochard, J. R. Rice, *J. Geophys. Res.* **105**, 25891 (2000).
18. G. G. Adams, *J. Appl. Mech.* **68**, 81 (2001).
19. J. Weertman, *J. Geophys. Res.* **85**, 1455 (1980).
20. D. J. Andrews, Y. Ben-Zion, *J. Geophys. Res.* **102**, 553 (1997).
21. Y. Ben-Zion, Y. Q. Huang, *J. Geophys. Res.* **107**, 10.1029/2001JB000254 (2002).
22. G. G. Adams, *J. Appl. Mech. Trans. ASME* **62**, 867 (1995).
23. G. G. Adams, *J. Appl. Mech. Trans. ASME* **65**, 470 (1998).
24. R. A. Harris, S. M. Day, *Bull. Seismol. Soc. Am.* **87**, 1267 (1997).
25. R. A. Harris, J. F. Dolan, R. Hartleb, S. M. Day, *Bull. Seismol. Soc. Am.* **92**, 245 (2002).
26. Bimaterial contrast is one of the two reasons cited in an attempt to explain bilateral rupture asymmetry for Izmit (3). However, as correctly pointed out by Andrews (29) and by Weertman (30), Bouchon *et al.* (3) assume that the higher wave speed material lies to the south of the fault, which is inconsistent with a westward growing +GR rupture. The present experiments conclusively settle the issue provided that the material to the north is the higher wave speed solid. However, if the fault is hypothesized to be homogeneous (3, 31), the rupture velocity asymmetry may alternatively be attributed to differences in fault plane morphology to the west (rough) and to the east (smooth) of the hypocenter.
27. E. Zor *et al.*, *Geophys. Res. Lett.* **30**, 10.1029/2003GL018129 (2003).
28. C. Thurber *et al.*, *Geophys. Res. Lett.* **30**, 10.1029/2002GL016004 (2003).
29. D. J. Andrews, *Geophys. Res. Lett.* **29**, 10.1029/2001GL014126 (2002).
30. J. Weertman, *Geophys. Res. Lett.* **29**, 10.1029/2001GL013916 (2002).
31. M. Bouchon, A. J. Rosakis, *Geophys. Res. Lett.* **29**, 10.1026/2002GL015096 (2002).
32. A.R., H.K., and K.X. acknowledge the support of NSF EAR 0207873 and Office of Naval Research N00014-03-0435. J.R. acknowledges the support by NSF EAR 0125702.

2 December 2004; accepted 16 February 2005
 10.1126/science.1108193

Enhanced Diapycnal Mixing by Salt Fingers in the Thermocline of the Tropical Atlantic

R. W. Schmitt,* J. R. Ledwell, E. T. Montgomery,
K. L. Polzin, J. M. Toole

Diapycnal mixing plays a significant role in the ocean's circulation and uptake of heat and carbon dioxide, but has not been quantified in salt finger–driven thermohaline staircases. We recently performed a tracer release experiment in the western tropical Atlantic staircase at ~400 m depth. The observed dispersion implies an effective diapycnal diffusivity for tracer and salt of 0.8 to $0.9 \times 10^{-4} \text{ m}^2/\text{s}$. Temperature microstructure data interpreted in terms of a vertical production-dissipation balance yields a smaller effective diffusivity for heat of $0.45 (\pm 0.2) \times 10^{-4} \text{ m}^2/\text{s}$, consistent with salt fingers and well above the mixing ascribable to mechanical turbulence.

Since the discovery of salt fingers by Stern (1) in 1960, there has been much speculation about their role in oceanic mixing (2, 3). Fieldwork has suggested the importance of salt fingers to the dynamics of fine scale intrusions (4) and in regional water mass transformations (5–7), but their contribution to the rates of diapycnal mixing has not been quantified. The difficulty has been due in part to uncertainties in the methods of measuring and interpreting oceanic microstructure for double-diffusive mixing. Because tracer release experiments (TREs) are complementary to the analysis of microstructure data, they offer an independent means of quantifying diapycnal mixing in the ocean. TREs have been successfully carried out in an eastern subtropical gyre area [the North Atlantic Tracer Release Experiment, NATRE (8, 9)] and an abyssal region above rough bathymetry (10, 11). To address the salt finger question, we recently performed a TRE in the western tropical North Atlantic where warm, high-salinity Subtropical Underwater overlies cooler, fresher Antarctic Intermediate Water (5, 12). Here the stratification through the main thermocline is often characterized by 10 or more layers (each 10 to 30 m thick), having near-uniform temperature and salinity, that are separated by 0.5- to 5-m-thick high-gradient interfaces (Fig. 1A). Such “staircase” layering in both the laboratory and ocean is a well-known manifestation of the salt finger form of double-diffusive convection (13), in which the greater diffusivity of heat over that of dissolved salts allows release of the potential energy of the salinity distribution in centimeter-scale convection cells.

We conducted fine structure and microstructure surveys in the western tropical Atlantic in January to February and October to December

2001 using the free-falling “high-resolution profiler” (HRP) (14). Temperature (T) and salinity (S) profiles from many sites revealed a step-like stratification through the main thermocline between 200 and 600 m depth (Fig. 1A). Temperature contrasts between adjacent layers sometimes approached 1°C . Staircase stratification was most commonly observed within the frontal region between the saltier North Atlantic Central Waters and fresher South Atlantic Central Waters, indicated in Fig. 2 by the salinity contour 35.1 on the density surface of the release. Our survey found a broad expanse

where staircase stratification may occur that was generally consistent with the area defined by prior surveys (5, 12).

The density ratio ($R_\rho = \alpha\Delta T/\beta\Delta S$, where the thermal expansion coefficient is $\alpha = \frac{-1}{\rho} \frac{\partial \rho}{\partial T}$, the haline contraction coefficient is $\beta = \frac{1}{\rho} \frac{\partial \rho}{\partial S}$, and ΔT and ΔS are the layer-to-layer temperature and salinity changes) was 1.4 to 1.7 across most interfaces, which indicates a strong propensity for salt fingering (15). Thermohaline staircases readily form in laboratory experiments with density ratios near one; salt fingers occupy the high-gradient interfaces and provide an up-gradient buoyancy flux that drives convective motions in the layers, maintaining weak stratification there. Thermal microstructure on the oceanic interfaces was elevated above levels normally found in the subtropical thermocline (even after accounting for the enhanced vertical temperature gradients of the interfaces). The dissipation rate of thermal variance ($\chi_\theta = 6\kappa_\theta \overline{\theta_z^2}$), where θ_z represents the microscale (centimeter) fluctuations in the vertical temperature gradient, shows prominent peaks at the interfaces (Fig. 1B), although evidence for the narrow-bandwidth temperature anomalies previously seen in horizontally towed data (16, 17) was lacking in these vertical profiles. A measure of the propensity for salt fingering is the theoretical salt finger growth rate that can be calculated from the separately measured meter-scale verti-

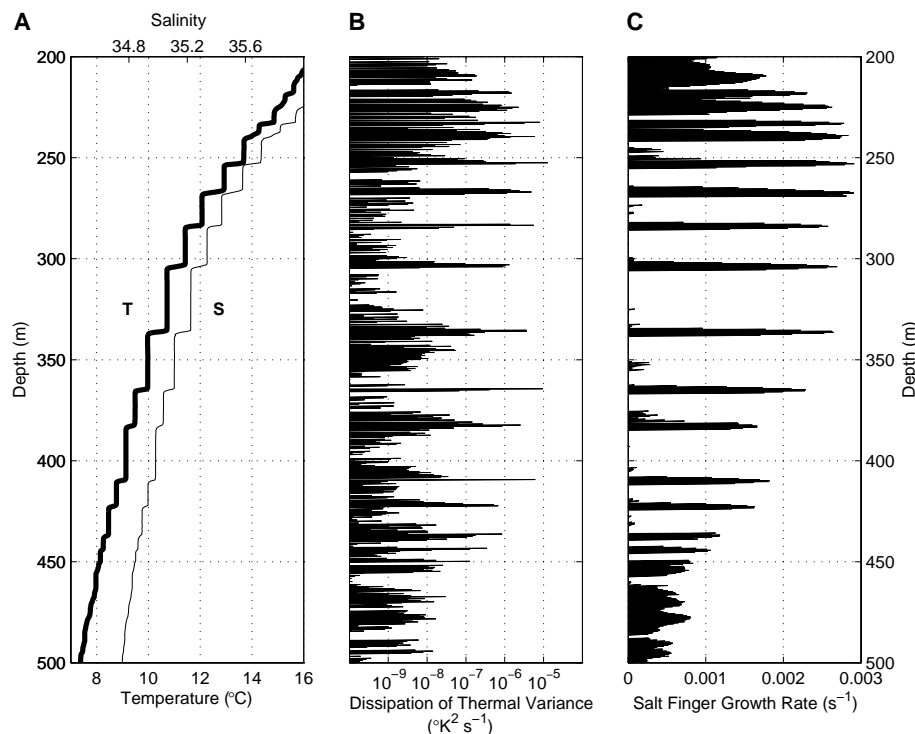


Fig. 1. Profiles of potential temperature and salinity (A), the dissipation rate of thermal variance (B), and the salt finger growth rate (C) for HRP dive 38 ($12^\circ 24.7'N$, $53^\circ 40.0'W$ on 7 November 2001). The SF_6 tracer was injected into the layer with potential temperature of about 10°C (centered on 350 db in this profile) in January 2001.

Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA.

*To whom correspondence should be addressed. E-mail: rschmitt@whoi.edu

cal temperature and salinity gradients (15) (Fig. 1C). The finger growth rate is large at the interfaces (steps) and zero in the mixed layers (risers), a pattern very similar to that of log (χ_θ). This result is consistent with the “frozen-growth” model of oceanic salt fingers (15, 18).

Salt fingers transport buoyancy-scaled T and S vertically at different rates, and so, the vertical salt flux cannot be directly assessed from thermal microstructure data. Nonetheless, we infer that the normalized vertical salt flux is greater than that for heat within the staircase by inspection of the horizontal variation of temperature and salinity within the layers. As observed previously (5), spatial variations of temperature and salinity within a given layer are highly correlated, with the ensemble of T/S values following a distinct slope across density surfaces at an apparent lateral density ratio of 0.85. This ratio is consistent with a vertical flux divergence due to salt fingering and is not consistent with any other known oceanic mixing process (19). To quantify the mixing rate for salt (and other scalars) in the staircase, we released sulfur hexafluoride (SF_6) in one of the layers and measured its subsequent dispersion. SF_6 should mix like salt in the ocean because the molecular diffusivities differ by less than 10%. The advective-diffusive balances are therefore nearly the same, and so the ratio of the vertical flux to vertical gradient is expected to be the same when quantities are averaged over a volume containing many mixing events.

During the January to February 2001 cruise, 175 kg of SF_6 was released into a midstaircase layer in a patch centered near

12°45'N, 53°45'W. The layer temperature was about 10°C and salinity was about 35.1, T/S values near the median of those sampled regionally for this layer. The tracer patch was laid out with a series of eight tows of an injector sled that pumped liquid SF_6 through atomizing nozzles. Sled depth was controlled by the vessel winch with guidance from a sled-mounted conductivity-temperature-depth (CTD) instrument and additional temperature-conductivity sensors placed 10 m above and below. Tracer was pumped only when the sled was positioned within the well-mixed layer and the layer potential density anomaly was between 27.041 kg/m³ and 27.052 kg/m³. This initial patch fell within an area that was ~28 km in diameter. A survey completed within 2 weeks of injection indicated that nearly 92% of the tracer was in the 10°C layer, some having already mixed to neighboring interfaces and layers.

Nearly 10 months later, in October to December 2001, the tracer patch was again sampled during an extensive regional survey (Fig. 2). At each station, the SF_6 concentration in discrete water samples collected in conjunction with CTD casts was measured using electron capture gas chromatography. Concentration estimates were then adjusted to remove a background profile estimated from data obtained outside the patch. The vertical integral of adjusted SF_6 concentration was mapped over an area of 1.3×10^6 km². Tracer dispersal was anisotropic (strongly zonal), with SF_6 found in three main patches. About 50% of the injected tracer could be

accounted for by this survey; subsequent survey work in 2003 suggested that the bulk of the “missing” tracer was in the Caribbean Sea west of the October to December 2001 survey area. The concentration profile data were subsequently integrated horizontally along isopycnals for the sampled regions east and west of the Lesser Antilles (Fig. 3). Over the 293 days between the midpoints of the injection and survey, the square root of the second moment of the vertical tracer distribution grew to 61 m. In contrast, this quantity grew to only 30 m after 12 months in the NATRE (8, 9). Although the normalized concentration profiles have similar shape, the center of mass of the eastern region’s distribution is about 20 m below that for the Caribbean region (and the injection isopycnal surface). This pattern appears to be partly the consequence of vertical shear in the North Equatorial Current that dominates the study region and partly the result of some sinking of the tracer.

As was described in (9), a model for the diapycnal dispersion was applied to the observations to diagnose the tracer’s effective diapycnal diffusivity. The resulting diffusivity estimate is 0.8 to 0.9×10^{-4} m²/s, with most of the uncertainty arising from the interpretation of the offset of the center of mass of the distribution. For reference, tracer diffusivity estimates from the eastern North Atlantic thermocline in NATRE were 0.12×10^{-4} m²/s for the first 6 months and 0.17×10^{-4} m²/s over 24 months. The present tracer-based estimate can be compared with the diffusivity inferred from the microstructure measurements. Several microstructure-based measures of mixing intensity are available based on the dissipation rate of thermal variance, χ_θ ; the dissipation rate of turbulent kinetic energy, ϵ ; and two different models for their interpretation: one for shear-driven turbulence and one for salt fingering. The formulae for diffusivities from the two models are as follows:

Turbulence (20, 21) (with flux Richardson number, $R_f = 0.17 \pm 0.03$):

$$K_\theta = K_S = K_p = \frac{R_f}{1 - R_f} \frac{\epsilon}{N^2} \approx 0.2 \frac{\epsilon}{N^2}$$

$$K_\theta = K_S = \frac{\chi_\theta}{2\theta_z^2}$$

Salt fingers (22) (with $R_p = 1.7$ and flux ratio $\gamma = 0.75$):

$$K_S = \frac{R_p - 1}{1 - \gamma} \frac{\epsilon}{N^2} \approx 2.8 \frac{\epsilon}{N^2}$$

$$K_S = \frac{R_p}{\gamma} \frac{\chi_\theta}{2\theta_z^2} \approx 2.3 K_\theta$$

$$K_\theta = \frac{\chi_\theta}{2\theta_z^2}$$

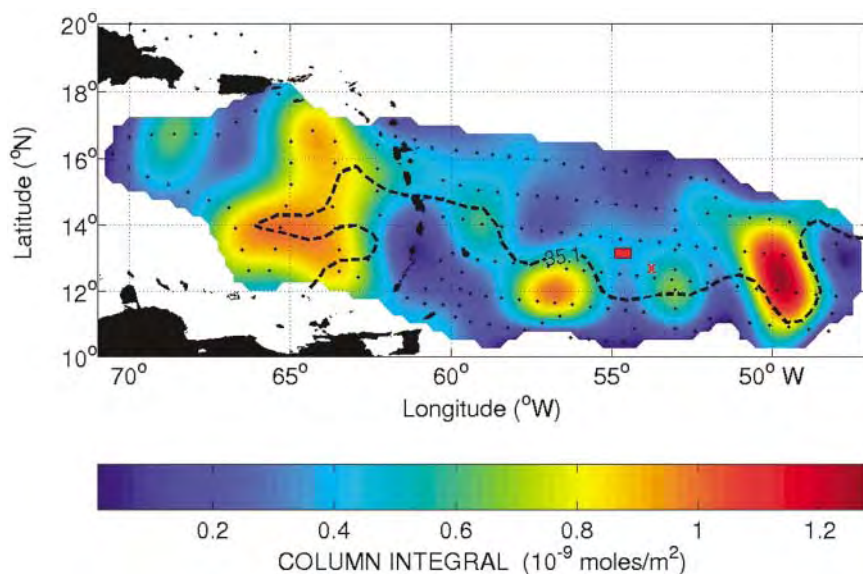


Fig. 2. Salt finger TRE map. The “X” near 54°W indicates the release area; the red rectangle northwest of the X shows the area where the tracer was found about 2 weeks after the release. The color map shows the column integral of tracer above background, in 10^{-9} mol/m², about 10 months after the release, with the stations shown as dots. The column integral at the stations north of the Greater Antilles was indistinguishable from background. The dashed lines indicate where the salinity was 35.1 on the density surface of the tracer release for the 10-month survey. On this surface, it is saltier to the north and fresher to the south. The tracer was not fully delimited, particularly in the northeast and the southwest.

In the above, K_θ , K_S , and K_ρ are the eddy diffusivities for heat, salt, and density; $N^2 = -\frac{g}{\rho} \frac{\partial \rho}{\partial z}$ is the squared buoyancy frequency, where g is gravity; and $\bar{\theta}_z$ is the mean vertical temperature gradient.

The key distinction between these two processes is found in the eddy diffusivities for heat and salt: they are equal for fully developed turbulence, but differ by a factor of ~ 2 for salt fingers, assuming a flux ratio of 0.75, a value consistent with laboratory, theory, and oceanic observations (23). From the HRP survey, overall averages of χ_θ , ϵ , and the vertical temperature and density gradients were formed between the potential density anomaly surfaces 26.968 and 27.128, which make up the 1-SD range of the vertical tracer spread. In this density interval $\langle \chi_\theta \rangle = 3.5 \times 10^{-8} \text{ K}^2/\text{s}$, $\langle \epsilon \rangle = 1.0 \times 10^{-9} \text{ W/kg}$, with 95% confidence intervals of 2.3 to $5.1 \times 10^{-8} \text{ K}^2/\text{s}$ and 0.4 to $2.3 \times 10^{-9} \text{ W/kg}$, respectively (24). These large-area average dissipations include both layers and interfaces and are somewhat larger than prior microstructure observations at one staircase site in the region (16, 25), which probably reflects the greater range of phenomena likely to be observed over this wide region. The foregoing formulae yield the diffusivities shown in Table 1 when applied to these overall averages. We find best agreement between K_S inferred from the tracer dispersion and that derived by applying the salt finger model to the thermal dissipation data. The diffusivity estimated from the conventional turbulence model applied to the viscous dissipation data is a factor of five too small to explain the observed tracer dispersion.

Parts of the foregoing analysis are based on conventional vertical production-dissipation balances for temperature microstructure (21). Lateral stirring on isopycnal surfaces would complicate this interpretation. Within a thermohaline front, temperature and salinity anomalies can be generated by double-diffusive fluxes that cause the anomalies to migrate across density surfaces (26–28) and may even provide a mechanism for staircase formation (29). Such migration may account for some of the tracer dispersion and would alter the interpretation of the thermal dissipation measurements (30, 31). Similarly, some fraction of the observed turbulent dissipation is likely to be contributed by internal wave processes. However, the enhanced diapycnal dispersion of SF_6 in the staircase region relative to NATRE, the strongly layered stratification with few temperature inversions, the distinct temperature-salinity relations within the layers (5), the enhanced thermal microstructure in the salt-finger-favorable depth intervals with small turbulent dissipation, and past evidence for bandwidth-limited microstructure in these steps (16, 17) lead us to conclude that salt fingering is contributing to enhanced vertical mixing in the western tropical thermocline of the North

Atlantic with an effective diapycnal diffusivity for salt and tracers twice that of heat.

The significant vertical dispersion of tracer observed in this thermohaline staircase supports the idea that salt fingers significantly enhance mixing in certain parts of the main thermocline. Our derived salt diffusivity of $0.8\text{--}0.9 \times 10^{-4} \text{ m}^2/\text{s}$ is an order of magnitude larger than that predicted for typical internal wave breaking within the mid-latitude thermocline. Indeed, for this low-latitude region, parameterization of mixing supported by the background internal wave field (32, 33) indicates that a diffusivity of only $\sim 0.02 \times 10^{-4} \text{ m}^2/\text{s}$ should be expected. The tracer-derived diffusivity is also larger than that implied by microstructure measurements pre-

viously made in this staircase (16, 25, 34). However, it is in agreement with the salt finger model applied to those dissipation data (35), as well as to our new observations. Notably, the diapycnal tracer mixing rate observed in the western tropical Atlantic is 5 times that observed in the eastern subtropical Atlantic during NATRE, because of the presence of the thermohaline staircase. The staircase appears to transform the T - S structure of the thermocline waters entering the Caribbean, increasing the salinity and density of Antarctic Intermediate Water (36) and preconditioning it for sinking at higher latitudes. The efficient vertical transport within this strong tropical thermocline must be taken into account in oceanic and climate models, where the param-

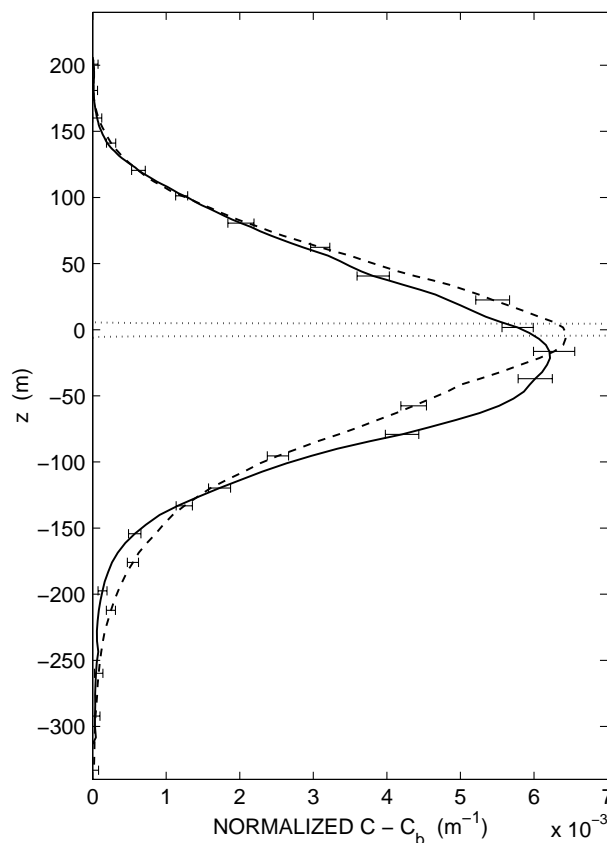


Fig. 3. Average vertical distribution of tracer about the injection density surface east of the Caribbean (solid line) and within the Caribbean Sea (dashed line) after 10 months. Tracer concentrations have been normalized such that the vertical integral is unity. The dotted lines delimit the initial tracer distribution. Averaging was done in density space, and the profiles were converted to physical space through the mean density profile for the stations east of Barbados.

Table 1. Diffusivities for salt and heat estimated from HRP measurements of χ_θ , the dissipation rate of thermal variance and ϵ , the dissipation rate of kinetic energy (penultimate row) $\times 10^4 \text{ m}^2/\text{s}$. The separate formulae for salt fingering and turbulent interpretations of the data are given. The estimates are based on the mean quantities over all stations occupied during the fall cruise between the potential density surfaces of 1026.968 and 1027.128 kg/m^3 , as these surfaces represent the 1-SD spread of the tracer. The 95% confidence interval for each estimate is given in the bottom row, based on statistics of the observed microstructure variables (24). Additional systematic errors may arise due to the uncertainties in the constants of the formulae.

Finger formulae (for $R_p = 1.71$, $\gamma = 0.75$)				Turbulence formulae (for $R_f = 0.17$)			
K_S		K_θ		K_S		K_θ	
$R_p \chi_\theta$	$R_p - 1 \epsilon$	$\frac{\chi_\theta}{2\bar{\theta}_z^2}$	$\frac{\gamma (R_p - 1) \epsilon}{R_p (1 - \gamma) N^2}$	$\frac{\chi_\theta}{2\bar{\theta}_z^2}$	$\frac{R_f \epsilon}{1 - R_f N^2}$	$\frac{\chi_\theta}{2\bar{\theta}_z^2}$	$\frac{R_f \epsilon}{1 - R_f N^2}$
1.03	2.40	0.45	1.07	0.45	0.17	0.45	0.17
0.56–1.8	0.8–6.6	0.26–0.75	0.36–2.7	0.26–0.75	0.06–0.42	0.26–0.75	0.06–0.42

eterization of diapycnal mixing continues to be a major uncertainty in assessing the ocean's ability to sequester heat, pollutants, and carbon dioxide.

References and Notes

1. M. E. Stern, *Tellus* **12**, 172 (1960).
2. R. B. Lambert, W. Sturges, *Deep-Sea Res.* **24**, 211 (1977).
3. R. W. Schmitt, *J. Phys. Oceanogr.* **11**, 1015 (1981).
4. B. Ruddick, *J. Phys. Oceanogr.* **22**, 1274 (1992).
5. R. W. Schmitt, H. Perkins, J. D. Boyd, M. C. Stalcup, *Deep-Sea Res.* **34**, 1697 (1987).
6. M. Tsuchiya, L. D. Talley, *J. Geophys. Res.* **103**, 12,899 (1998).
7. A. P. S. Wong, G. C. Johnson, *J. Phys. Oceanogr.* **33**, 1493 (2003).
8. J. R. Ledwell, A. J. Watson, C. S. Law, *Nature* **364**, 701 (1993).
9. J. R. Ledwell, A. J. Watson, C. S. Law, *J. Geophys. Res.* **103**, 21,499 (1998).
10. K. L. Polzin, J. M. Toole, J. R. Ledwell, R. W. Schmitt, *Science* **276**, 93 (1997).
11. J. R. Ledwell, E. T. Montgomery, K. L. Polzin, R. W. Schmitt, J. M. Toole, *Nature* **403**, 179 (2000).
12. J. D. Boyd, H. Perkins, *Deep-Sea Res.* **34**, 337 (1987).
13. R. W. Schmitt, *Annu. Rev. Fluid Mech.* **26**, 255 (1994).

14. R. W. Schmitt, J. M. Toole, R. L. Koehler, E. C. Mellinger, K. W. Doherty, *J. Atmos. Ocean. Tech.* **5**, 484 (1988).
15. R. W. Schmitt, *Deep-Sea Res.* **26A**, 23 (1979).
16. R. Lueck, *Deep-Sea Res.* **34**, 1677 (1987).
17. G. Marmorino, W. K. Brown, W. D. Morris, *Deep-Sea Res.* **34**, 1667 (1987).
18. A. E. Gargett, R. W. Schmitt, *J. Geophys. Res.* **87**, 8017 (1982).
19. T. J. McDougall, *Deep-Sea Res.* **38**, 367 (1991).
20. T. R. Osborn, *J. Phys. Oceanogr.* **10**, 83 (1980).
21. T. Osborn, C. Cox, *Geophys. Fluid Dyn.* **3**, 321 (1972).
22. T. J. McDougall, in *Small-Scale Turbulence and Mixing in the Ocean*, J. N. a. B. Jamart, Ed. (Elsevier Oceanography Ser., vol. 46, Elsevier, New York, 1988), pp. 21–36.
23. L. St. Laurent, R. W. Schmitt, *J. Phys. Oceanogr.* **29**, 1404 (1999).
24. The 95% confidence limits for the diffusivities were calculated using a "bootstrap" method with the statistics of 1000 random subsamples of half of the 165 stations for each of the measured variables. This treats each station as independent but does not account for possible systematic errors in the flux ratio or flux Richardson number.
25. M. C. Gregg, T. Sanford, *Deep-Sea Res.* **34**, 1689 (1987).
26. M. E. Stern, *Deep-Sea Res.* **14**, 747 (1967).
27. J. M. Toole, D. T. Georgi, *Prog. Oceanogr.* **10**, 123 (1981).
28. C. J. Garrett, *J. Phys. Oceanogr.* **12**, 952 (1982).

29. W. J. Merryfield, *J. Phys. Oceanogr.* **30**, 1046 (2000).
30. T. M. Joyce, *J. Phys. Oceanogr.* **7**, 626 (1977).
31. R. Ferrari, K. Polzin, *J. Phys. Oceanogr.* (in press).
32. M. C. Gregg, *J. Geophys. Res.* **94**, 9686 (1989).
33. K. Polzin, J. M. Toole, R. W. Schmitt, *J. Phys. Oceanogr.* **25**, 306 (1995).
34. E. Kunze, in *Double-Diffusive Convection*, A. Brandt and J. Fernando, Ed. *Geophys. Monogr.* **94**, 313 (1995).
35. R. W. Schmitt, in *Small-Scale Turbulence and Mixing in the Ocean*, J. N. a. B. Jamart, Ed. (Elsevier Oceanography Ser., vol. 46, Elsevier, New York, 1988), pp. 435–452.
36. W. J. Schmitz, J. R. Luyten, R. W. Schmitt, *Bull. Mar. Sci.* **53**, 1048 (1993).
37. We wish to thank the captains and crews of the *R/Vs Oceanus* and *Seward Johnson* for consistently fine work throughout our cruises. We also thank D. Wellwood, T. Bolmer, T. Farrar, T. Donoghue, B. Guest, C. Sellers, S. Birdwhistell, S. Sutherland, A. deBoer, L. Houghton, S. Ledwell, A. Benitez, G. Hernandez, and R. Brathwaite for scientific assistance during the cruises. This work was supported by the National Science Foundation under grants OCE-0081502 and OCE-0350743. This is Contribution Number 11284 of the Woods Hole Oceanographic Institution.

14 December 2004; accepted 16 March 2005
10.1126/science.1108678

Insect-Resistant GM Rice in Farmers' Fields: Assessing Productivity and Health Effects in China

Jikun Huang,^{1*} Ruifa Hu,¹ Scott Rozelle,² Carl Pray³

Although no country to date has released a major genetically modified (GM) food grain crop, China is on the threshold of commercializing GM rice. This paper studies two of the four GM varieties that are now in farm-level preproduction trials, the last step before commercialization. Farm surveys of randomly selected farm households that are cultivating the insect-resistant GM rice varieties, without the aid of experimental station technicians, demonstrate that when compared with households cultivating non-GM rice, small and poor farm households benefit from adopting GM rice by both higher crop yields and reduced use of pesticides, which also contribute to improved health.

Despite promises that GM crops could make a contribution to the reduction of hunger throughout the world, GM varieties are primarily used for industrial crops, such as cotton, and feed crops for animals (1–3). The difficulties of commercializing GM rice (and other food crops) appear to be causing declines in the amount and direction of public and private biotechnology research (4). Consequently, GM rice has not been commercialized anywhere in the world, and little is in the pipeline in most countries. Even China, a country

that aggressively commercialized Bt cotton and invested heavily into research on GM food crops, has not commercialized any major food crops.

One reason that commercialization may not have proceeded is that there has been little independent evidence on whether GM food crops would really improve farmer welfare. This study's objective is to report on the results of an economic analysis that uses data from eight rice preproduction trial sites in China. We attempt to answer three questions: Does GM rice help reduce pesticide use in the fields of farmers? Do the new varieties of GM rice increase the yields for farmers? Are there any identifiable health effects on the farmers that adopt GM rice strains?

China's biotechnology research program has generated a wide array of new technologies, including several GM rice varieties (5). A number of GM rice varieties have entered and passed field and environmental release trials, and four varieties are in preproduction

trials in farmers' fields. Two of the varieties—the two in which the scientists that developed the varieties gave our study team permission to undertake economic analysis—are the focus of this study (5). One variety, GM Xianyou 63, was created to be resistant to rice stem borer and leaf roller by insertion of a Chinese-created *Bacillus thuringiensis* (Bt) gene (5, 6). The other variety, GM II–Youming 86, also was created to be resistant to rice stem borers, but in this case, the resistance was created by introducing a modified cowpea trypsin inhibitor (CpTI) gene into rice (5–8). The insect-resistant GM varieties entered preproduction trials in 2001.

The nature of China's preproduction trial system has facilitated the analysis of the effect of insect-resistant GM rice on farm households before commercialization. The preproduction trials of GM Xianyou 63 are being conducted by farmers in seven villages in five counties in Hubei province. The trials for GM II–Youming 86 are being conducted in one village in Fujian province. In the preproduction villages, households were randomly selected to participate in the study. All of the farmers that were randomly selected did participate (i.e., there were no drop-outs), and so all farmers in the sample villages can be divided into two groups—adopters and nonadopters. Each adopter was provided with a fixed amount of insect-resistant GM rice seed. For households with limited land size, the seed was enough to cover all of their plots (henceforth, full adopters). Others received only enough to cover part of their plots (partial adopters). Except for being provided insect-resistant GM rice seed (at the same price as they would have paid for non-GM varieties), there were no subsidies, and adopters cultivated the insect-resistant GM rice without the assistance of technicians. Because farmers use their own periodic, in-field

¹Center for Chinese Agricultural Policy, Institute of Geographical Sciences and Natural Resource Research, Chinese Academy of Sciences, Jia 11, Datun Road, Beijing 100101, China. ²Department of Agricultural and Resource Economics, University of California, 1 Shields Avenue, Davis, CA 95616, USA. ³Department of Agricultural, Food, and Resource Economics, Rutgers University, 55 Dudley Road, New Brunswick, NJ 08901–8520, USA.

*To whom correspondence should be addressed. E-mail: jkhuang.ccap@igsnr.ac.cn

observations on the severity of pest infestation to decide whether or not to apply pesticides on both the insect-resistant GM and non-GM rice (that is, they are not following a prescribed dosage), the study can provide an estimate of the amount of farm-level pesticide reduction that can come from the adoption of the insect-resistant GM rice.

Our analysis presented here is based on surveys of a randomly selected subsample of households in the preproduction villages. During the first year of the study (2002), in six of the eight sample villages, there were only a limited number of adopters, and so all of them were chosen (some were full; the rest were partial adopters). A similar number were randomly chosen from all adopters in the other two villages. In total, 40 adopters (28 partial and 12 full) were chosen in 2002. In addition, 37 nonadopters (about one for each adopter) were chosen randomly from the pool of non-adopters in each village. In total, 77 households were surveyed in 2002. During 2003, a similar strategy was used, but because more insect-resistant GM seed was distributed, more adopters were added to the survey. Overall, 101 were interviewed in 2003 (32 nonadopters, 53 partial adopters, and 16 full adopters). There were 69 households that were interviewed in both years.

The enumerators, using producer-recall interviewing techniques, collected information on inputs and outputs for all of the plots on which the farmers produced rice, including detailed information on pesticide use and the variety of rice grown. Farmers also recounted the prices paid for pesticides and whether or not the plot was adversely affected by a weather shock. In total, the survey obtained data from 347 rice production plots: 123 plots planted with the insect-resistant GM rice varieties and 224 plots planted with non-GM rice.

Data from the surveys demonstrate that the characteristics of rice producers using the insect-resistant GM rice and non-GM rice are nearly identical and that the main difference between the households is in the level of pesticide use (9). For example, there is no statistical difference between the size of the farm or the plot or plots, the share of rice in the household's cropping pattern, or the household head's age or education. In contrast, there is a large difference in the use of pesticides (Table 1) (10). GM rice farmers apply the same types of pesticides but apply them less than once per season (0.5 times) compared with 3.7 times per season by non-GM rice farmers. The difference in the levels of pesticide use on insect-resistant GM and non-GM rice is statistically significant. On a per hectare basis, the quantity of and expenditure on pesticides of non-GM rice production is 8 to 10 times as high, respectively, as those for insect-resistant GM rice. Insect-resistant GM rice adopters

spend only 31 yuan per season per hectare on only 2.0 kg of pesticide for spraying for pests, whereas nonadopters spend 243 yuan for 21.2 kg.

Because other factors might affect pesticide use when comparing insect-resistant GM rice and non-GM rice, multiple regression can determine the net impact of the adoption of insect-resistant GM varieties on pesticide use. To estimate a use function for pesticide by China's rice farmers in the sample areas, the following model is used:

$$\text{Pesticide use} = f(\text{GM rice varieties, pesticide price, weather effects, year effects, producer and farm characteristics}) \quad (1)$$

Equation (1) is similar to models that have been used elsewhere in the literature (11, 12). To empirically estimate Eq. (1), the data from the survey are used to create variables that are based on standard definitions (13). The dependent variable for the analysis is the quantity of pesticides used per season (although substantively identical results are generated from either the number of sprayings per season or the value of pesticide use). The independent variable of interest, the use of the insect-resistant GM rice varieties, is measured by including a single dummy variable (GM rice, both varieties) which equals 1 if the farmer used either GM Xianyou 63 or GM II-Youming 86. In an alternative specification, the use of GM rice is measured by including two GM variety-specific dummy variables (GM Xianyou 63 and GM II-Youming 86) and two non-GM variety dummy variables. A set of household 0 to 1 indicator variables (108 of them—one for each sample household minus 1) is included to isolate the effect of GM varieties on pesticide use from observed and unobserved producer characteristics.

The regression analysis illustrates the importance of insect-resistant GM rice varieties in reducing pesticide use (Table 2, rows 2 to 6). The significant, negative coefficient on the "GM rice, both variety" variable means that GM rice use allows farmers to reduce pesticide use by 16.77 kg/ha, a reduction of nearly 80% (when compared with pesticide use of farmers using non-GM varieties—Table 1, row 3). The negative and significant coefficients on the GM Xianyou 63 and GM II-Youming 86 variables also demonstrate that each variety significantly reduces pesticides. Although the magnitudes of the coefficients differ, tests show that there is no statistical difference between the actual effects of the two insect-resistant GM varieties on pesticide use (Table 2, rows 3 and 4) (14).

The data also show that there is a difference, albeit narrower, between yields of insect-resistant GM and non-GM varieties.

According to the descriptive data in Table 1, the mean of insect-resistant GM rice yields (6364 kg/ha) is higher than those of non-GM varieties (6151), although only by 3.5%. A box plot also shows that the median of insect-resistant GM rice yields is marginally higher than those of non-GM rice (fig. S1). ANOVA tests that differentiate among year, village, and GM versus non-GM effects demonstrate that the effect is statistically significant (15).

Multiple regression analysis largely supports the descriptive results (Table 2). Holding all household-level effects, plot-specific inputs, and certain other plot characteristics constant, the yields of insect-resistant GM varieties are 6% higher than those of non-GM varieties. When examining the effects of specific varieties (compared with other conventional varieties—the base category), the yields of GM Xianyou 63 are shown to be 9% higher (at the 10% level of significance) than other conventional varieties. Although the yields of GM II-Youming 86 are not found to be significantly different from conventional non-GM varieties, this result in part may be due to the fact that there are relatively few observations (because preproduction trials of GM II-Youming 86 are from one village only, and there are relatively few farm households that were partial adopters). Therefore, according to the descriptive and multiple regression analyses, although the evidence on effect of the insect-resistant GM rice varieties on increasing yields is not as overwhelming as that which examines the relationship between the GM rice varieties and pesticides, the GM Xianyou 63 rice variety does appear to increase yields (between 6 and 9%) (16).

The high incidence of pesticide-related illness in households in developing countries, including China, created an interest in tracking the health effects of insect-resistant GM rice adoption (11, 12, 17). To assess the effects in this study's sample, enumerators asked

Table 1. Pesticide use and yields of insect-resistant GM rice adopters and nonadopters in preproduction trials in China, 2002–2003 (means ± SD). Insect-resistant GM rice includes two varieties, GM Xianyou 63 and GM II-Youming 86. Data are from the authors' survey.

Parameter	Adopters	Nonadopters
Pesticide spray (times)	0.50 ± 0.81	3.70 ± 1.91
Expenditure on pesticide (yuan/ha)	31 ± 49	243 ± 185
Pesticide use (kg/ha)	2.0 ± 2.8	21.2 ± 15.6
Pesticide spray labor (days/ha)	0.73 ± 1.50	9.10 ± 7.73
Rice yield (kg/ha)	6364 ± 1294	6151 ± 1517
No. of observations (plots)	123	224

households about how the use of pesticides affected their health during, or immediately after, the time that they applied pesticides (18). Specifically, the questionnaire asked the farmers, "During or after spraying for pesticides on your farm, did you suffer from any of the following symptoms: headaches, nausea, skin irritation, digestive discomfort, or other problems?" If the respondent answered "yes," a follow-up question was asked: "After beginning to feel poorly, did you take any one of the following actions: 1) visit a doctor; 2) go home and recover at home; 3) take some other explicit action to mitigate the symptoms?" If the respondent answered "yes" to both of the questions, it was recorded as a case of pesticide-induced illness.

In the same way that research on Bt cotton adoption showed that the productivity effects of Bt cotton were supplemented by positive

health effects (3), according to the analysis based on the survey data, similar effects occur within the sample rice-growing households. Among the sample farmers, there were no full adopters that reported being affected adversely by pesticide use in either 2002 or 2003 (Table 3). Of those that cultivated both insect-resistant GM and non-GM plots, 7.7% of households in 2002 and 10.9% of households in 2003 reported that their health was affected adversely by pesticide use; none, however, reported being affected after working on the sample GM plot. Of those that used only non-GM varieties, the health of 8.3% of households in 2002 and 3% in 2003 was affected adversely.

This study provides evidence that there are positive impacts of the insect-resistant GM rice on productivity and farmer health. Insect-resistant GM rice yields were 6 to 9% higher

than conventional varieties, with an 80% reduction in pesticide usage and a reduction in their adverse health effects. Such high potential benefits suggest that products from China's plant biotechnology industry could be an effective way to increase both competitiveness internationally and rural incomes domestically. The benefits are only magnified if the health effects are added. The implications of the commercialization of GM rice in China also could far exceed the productivity and health effects on its own producers. Paarlberg suggests that if China were to commercialize a major crop, such as rice, it is possible that it would influence the decisions about the commercialization of GM crops in the rest of the world (4).

Table 2. Estimated parameters using a household fixed-effects model for estimating the effect of insect-resistant GM rice varieties on farmers' pesticide application and the yields of households in preproduction trials in China. The coefficients from the multiple regression model represent the net effect of insect-resistant GM rice varieties on pesticide use and yield, with the other plot-varying variables in the model held constant. For rice variety dummies, the base value is other non-GM varieties. Model 1 has both varieties as one variable; model 2 has treated the two varieties separately. The use of household fixed effects is accomplished by including 108 household dummy variables (equals 1 for the household and 0 otherwise), which allows for the control for all unobserved non-time-varying producer and farm characteristics. Values are means \pm SD. The symbols *, †, and ‡ denote significance at 1, 5, and 10%, respectively. Data are from the authors' survey.

Variables	Pesticide use (kg/ha)		Yields (kg/ha) in log	
	Model 1	Model 2	Model 1	Model 2
Intercept	19.93 \pm 1.17*	19.78 \pm 1.32*	7.55 \pm 0.50*	7.61 \pm 0.51*
Variety dummies				
GM rice, both varieties	-16.77 \pm 1.28*		0.06 \pm 0.03‡	
Variety-specific dummy variables				
GM Xianyou 63		-17.15 \pm 2.60*		0.09 \pm 0.05‡
GM II-Youming 86		-25.33 \pm 5.48*		0.02 \pm 0.10
Non-GM Xianyou 63		1.04 \pm 2.61		-0.03 \pm 0.05
Non-GM II-Youming 86		-1.25 \pm 3.82		0.07 \pm 0.07
Control variables				
Pesticide price (yuan/kg)	-0.02 \pm 0.03	-0.02 \pm 0.03		
Natural disaster dummy (affected = 1)	8.56 \pm 2.65*	8.65 \pm 2.65*	-0.51 \pm 0.05*	-0.51 \pm 0.05*
2003 year dummy	-0.17 \pm 1.20	-0.01 \pm 1.24	-0.05 \pm 0.02†	-0.05 \pm 0.02†
Labor (log)			0.17 \pm 0.07†	0.17 \pm 0.07†
Fertilizer (log)			0.04 \pm 0.06	0.03 \pm 0.06
Machine (log)			0.00 \pm 0.01	0.00 \pm 0.01
Other inputs (log)			0.03 \pm 0.04	0.02 \pm 0.04
Pesticides (log)			0.00 \pm 0.00	0.00 \pm 0.00
Household dummy variables		Included but not reported		
No. of observations	347	347	347	347

Table 3. The effect of insect-resistant GM rice use on the health effects of farmers in sample preproduction village sites in China, 2002–2003. Full adopters planted insect-resistant GM rice only; partial adopters planted both GM and non-GM rice; and nonadopters planted non-GM rice only. The numbers are the percentage of sample households that were adversely affected by pesticides. Data are from the authors' survey.

Adverse health effects reported and year	Full adopters	Partial adopters		Nonadopters
		GM plot	Non-GM plot	
2002	0.0	0.0	7.7	8.3
2003	0.0	0.0	10.9	3.0

References and Notes

1. M. Qaim, D. Zilberman, *Science* **299**, 900 (2001).
2. It should be noted that despite there being no genetically modified major food grains being used anywhere in the world, there are more minor food crops, such as papaya and squash.
3. J. Huang, S. D. Rozelle, C. E. Pray, Q. Wang, *Science* **295**, 674 (2002).
4. R. Paarlberg, *Issues Sci. Technol. Online* (Spring 2003); available at www.issues.org/issues/19.3/paarlberg.htm.
5. See Supporting Online Material (SOM), section 1 on *Science Online*.
6. J. Tu *et al.*, *Nat. Biotechnol.* **18**, 1101 (2000).
7. C. Deng, G. Song, J. Xu, Z. Zhu, *Acta Bot. Sin.* **45**, 1084 (2003).
8. This paper discusses the line and the gene construct for CpTI and transfer of the gene into tobacco. Information on the transfer of the gene into rice is in preparation for publication (J. Huang, R. Hu, S. Rozelle, C. Pray).
9. See SOM, section 2. Insect-resistant GM rice strains are produced only to deal with pests and not to increase flavor or alter nutrition. As a consequence, there is no difference in prices between the insect-resistant GM rice and non-GM rice.
10. See SOM, section 7.
11. J. M. Antle, P. L. Pingali, *Am. J. Agric. Econ.* **76**, 418 (1994).
12. J. Huang, F. Qiao, L. Zhang, S. Rozelle, "Farm pesticide, rice production, and the environment" [Economy and Environment Program for Southeast Asia (EEPSEA) Res. Rep. 2001-RR3, International Development Research Center (IDRC), Singapore, 2001].
13. See SOM, section 3.
14. See SOM, section 4.
15. See SOM, section 5.
16. See SOM, section 6.
17. J. Huang, R. Hu, C. Pray, F. Qiao, S. Rozelle, *Agric. Econ.* **29**, 55 (2003).
18. See SOM, Section 8, for original and translation of questions.
19. We are grateful to Q. Zhang and Z. Zhu and their colleagues who developed GM rice in China for technical inputs and to P. J. Hines and three anonymous referees for helpful comments. The authors acknowledge the support of the National Natural Science Foundation of China (grants 70021001, 70333001, and 70325003) and the Chinese Academy of Science (KZCX3-SW-419).

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5722/688/DC1
 Materials and Methods
 SOM Text
 Fig. S1
 Tables S1 to S6
 References and Notes

21 December 2004; accepted 9 February 2005
 10.1126/science.1108972

A Rapid Shift in a Classic Clinal Pattern in *Drosophila* Reflecting Climate Change

P. A. Umina,¹ A. R. Weeks,² M. R. Kearney,²
S. W. McKechnie,¹ A. A. Hoffmann^{2*}

Geographical clines in genetic polymorphisms are widely used as evidence of climatic selection and are expected to shift with climate change. We show that the classic latitudinal cline in the alcohol dehydrogenase polymorphism of *Drosophila melanogaster* has shifted over 20 years in eastern coastal Australia. Southern high-latitude populations now have the genetic constitution of more northerly populations, equivalent to a shift of 4° in latitude. A similar shift was detected for a genetically independent inversion polymorphism, whereas two other linked polymorphisms exhibiting weaker clinal patterns have remained relatively stable. These genetic changes are likely to reflect increasingly warmer and drier conditions and may serve as sensitive biomarkers for climate change.

Clinal variation in genetic polymorphisms and traits occurs in natural populations of numerous species. Such variation provides strong evidence of natural selection and adaptation to different environmental variables, particularly climatic conditions (1). Perhaps the most compelling evidence for climatic selection generating clinal patterns comes from the many studies on latitudinal clinal variation in *Drosophila melanogaster*. This cosmopolitan species is closely associated with human ac-

tivities, and its abundance is typically high in farmed areas and in decaying fruit and vegetable matter. Latitudinal clines have been shown in *D. melanogaster* for traits such as body size, development time, and thermal resistance (2–4) and for various genetic markers at the allele and karyotype level (5–8).

The alcohol dehydrogenase (*Adh*) locus in this species is one of the most thoroughly studied examples of a genetic latitudinal cline in any organism, with the *Adh^S* allele increasing in frequency with decreasing latitude in both the Northern and Southern hemispheres (5, 9, 10). Temperature-related factors influence population allele frequencies both in the field (6, 11) and in the laboratory (12, 13); rainfall and humidity levels have also been implicated (5).

Using molecular markers for *Adh* (14), we recharacterized the *D. melanogaster* cline along coastal eastern Australia in 2002 and 2004 and compared the results to data from earlier studies (5, 15) that involved collections in 1979 and 1982. The combined data set from these different periods showed a significant difference in elevation but not in the slope of the regression lines (Fig. 1 and Table 1), with elevation shifting in the direction expected with rising temperatures and global warming (6, 16). Because the *Adh^S* allele that predominates in tropical populations is now at a higher frequency in all populations, there has been a tendency for populations to evolve toward the genetic composition of lower latitude populations. Conversely, we found no differences in either line slope ($t = 1.164$, $df = 30$, $P = 0.254$) or line elevation ($t = 0.842$, $df = 31$, $P = 0.406$) between the 2002 and 2004 collections. Nor did we find any differences in slope ($t = 0.973$, $df = 42$, $P = 0.336$) or elevation ($t = 0.305$, $df = 43$, $P = 0.762$) between the 1979 and 1982 collections.

In studies of latitudinal transects, the *Adh^S* allele has been positively associated with *In(2L)t*, a cosmopolitan inversion that also clines latitudinally (6, 7). *Adh* is located just outside the proximal breakpoint of this inversion, whereas α -*Gpdh*, the gene encoding another commonly studied cytosolic enzyme (*sn*-glycerol-3-phosphate dehydrogenase), is inside it; in natural populations, this locus is usually in disequilibrium with the inversion. We examined changes in patterns at both the α -*Gpdh* locus (14) and for *In(2L)t*, again pooling across years on the basis of nonsignificant changes in slope and elevation. In the earlier studies (7, 15), *In(2L)t* had been scored cytologically after laboratory culture of lines, but we used a molecular marker to

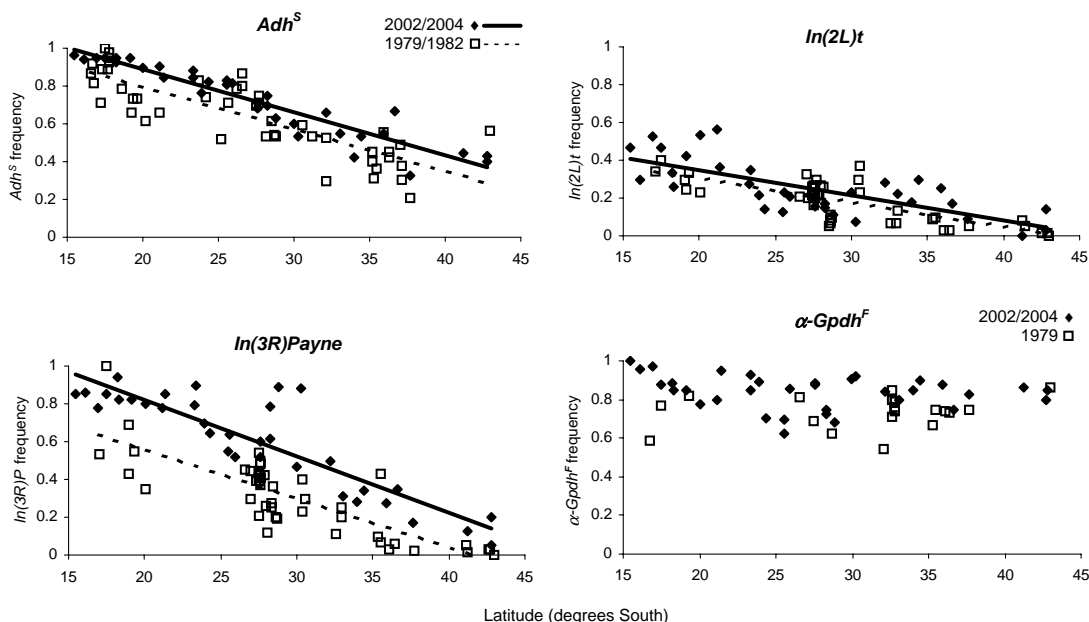
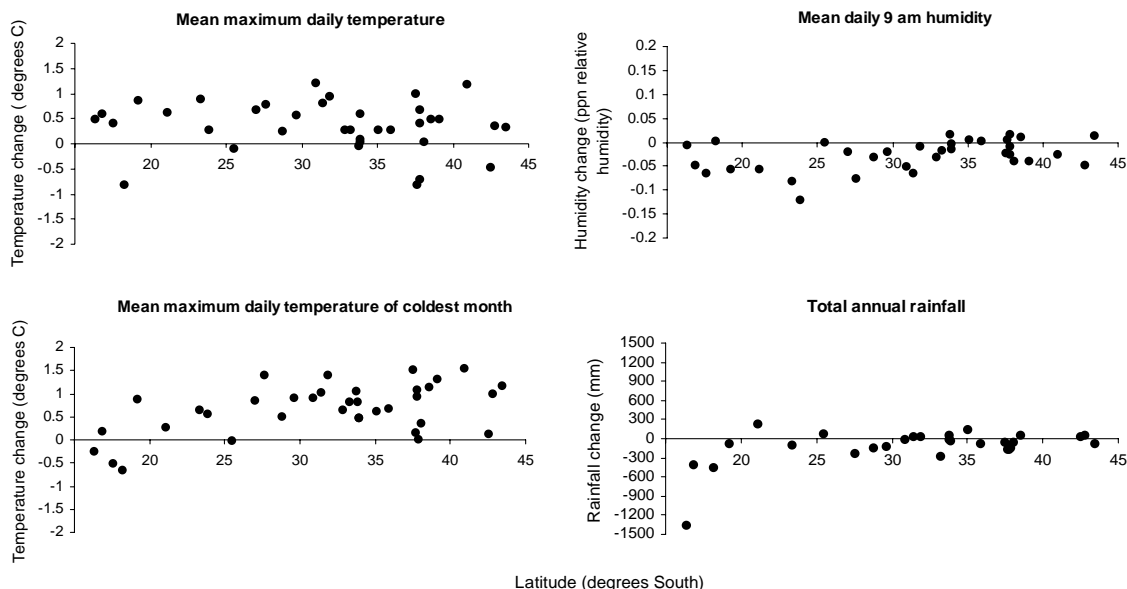


Fig. 1. Change in latitudinal patterns, between 1979 and 1982 and between 2002 and 2004, of the frequencies of the S allele of alcohol dehydrogenase (*Adh^S*), the inversions *In(3R)Payne* and *In(2L)t*, and the F allele of *sn*-glycerol-3-phosphate dehydrogenase (α -*Gpdh^F*). Open symbols and dashed lines indicate 1979–1982 pooled data points (except α -*Gpdh^F* where only 1979 data were available); solid symbols and solid lines indicate 2002–2004 pooled data.

Table 1. Linear regressions examining associations between markers and latitude in 1979–1982 and 2002–2004 collections of *D. melanogaster* from the east coast of Australia. Probability tests for differences between slopes and elevations of the regression equations are also provided.

Marker	Year	Regression equation	R^2	t	P	P values for comparison of two collections	
						Slope	Elevation
<i>Adh^S</i>	2002–2004	$y = -0.028x + 1.794$	0.899	16.873	<0.001	0.677	<0.001
	1979–1982	$y = -0.026x + 1.651$	0.664	9.314	<0.001		
<i>In(2L)t</i>	2002–2004	$y = -0.018x + 0.982$	0.551	6.272	<0.001	0.480	0.079
	1979–1982	$y = -0.020x + 1.009$	0.652	9.077	<0.001		
<i>In(3R)Payne</i>	2002–2004	$y = -0.033x + 1.797$	0.772	10.413	<0.001	0.586	<0.001
	1979–1982	$y = -0.035x + 1.619$	0.696	10.047	<0.001		

Fig. 2. Differences between 1979–1982 and 2002–2004 values for four climatic variables plotted against latitude. Positive values indicate higher scores for 2002–2004.



score this inversion directly on field flies, following Andolfatto *et al.* (14, 17). The 1979–1982 and 2002–2004 data for *In(2L)t* did not differ significantly in terms of either the slope or elevation of the regression lines (Fig. 1 and Table 1), although there was a nonsignificant trend in the same direction to that observed for *Adh*. Although a weak latitudinal cline has been suggested previously for α -*Gpdh* in Australia (5, 18), we found no significant linear association between α -*Gpdh* and latitude in the 2002 ($t = 0.945$, $P = 0.361$) and 2004 ($t = 1.675$, $P = 0.113$) collections (Fig. 1). Furthermore, there was no significant difference in the mean frequency of α -*Gpdh* between the 1979 and the pooled 2002–2004 populations ($t = -1.764$, $df = 53$, $P = 0.084$). Therefore, despite the genetic and possible adaptive interdependence of α -*Gpdh* and *In(2L)t* with *Adh*, they have remained relatively unchanged in latitudinal position.

We estimated the magnitude of the shift in elevation of the clinal pattern for the *Adh* polymorphism, as well as for a common cosmopolitan inversion on a different chromosome, *In(3R)Payne*, which also clines with latitude and has shifted in elevation (but not slope) in the past 20 years (19) (Fig. 1 and Table 1). For *Adh*, the shift corresponded to 3.9° in latitude

[bootstrap 95% confidence limits (CLs) of 2.01 and 5.92], equivalent to more than 400 km. For *In(3R)Payne* the shift was larger, at 7.3° in latitude (CLs of 4.85 and 10.51), corresponding to a shift of more than 800 km.

What factors might be responsible for these latitudinal shifts? *Adh* frequencies have previously been related to temperature and rainfall (5, 11, 16), and because these factors vary clinally, they are likely contenders. The climate along the eastern coast of Australia is gradually becoming warmer and drier, and recent evidence shows that over the past 50 years these changes have increased markedly as a result of human activities (20, 21). Mean temperature is increasing at most coastal locations at a rate of 0.1° to 0.3°C every 10 years, whereas rainfall is decreasing at a rate of 10 to 70 mm per year. These changes are also evident in a comparison of data from weather stations along the eastern coast of Australia covering the area sampled for the molecular markers. Average differences between 1978 and 1981 and between 2000 and 2003 (1 to 2 years before each set of collections) were compared for 36 weather stations along the eastern coast at low altitudes (<50 m) on the basis of data from the Australian Bureau of Meteorology (22). There were significant

differences (using paired t tests) between years in 11 of the 15 climatic variables considered. When data for four of these variables—change in mean daily maximum temperature of the coldest month ($t = 4.288$, $df = 35$, $P < 0.001$), change in average daily maximum temperature ($t = 7.218$, $df = 35$, $P < 0.001$), change in mean daily humidity at 9 a.m. ($t = -4.918$, $df = 29$, $P < 0.001$), and change in total annual rainfall ($t = -2.437$, $df = 28$, $P = 0.021$)—are plotted against latitude, they indicate higher mean daily maximum temperatures, lower humidity, and a decrease in rainfall at most latitudes (Fig. 2).

Combinations of climatic variables are likely to provide the best indicators of the shifts in genetic constitution, because changes in any one of these single climatic variables do not appear to account for the extent of the latitudinal shift observed. For instance, the 0.66°C shift in the maximum daily temperature of the coldest month across samples amounts to a shift in latitude of only 1.3°, which is outside the lower bootstrap confidence intervals of the shifts estimated for both *Adh* and *In(3R)Payne*. The combined effect of different climatic factors may underlie the changed environment that has driven the coastal shift in *Adh^S* and *In(3R)Payne* frequencies.

Climatically driven changes in the genetic constitution of *D. melanogaster* populations are not entirely unexpected, because climate-associated changes in chromosomal inversions have been documented for the *Drosophila* species *D. robusta* (23) and *D. subobscura* (24). In addition, previous *Drosophila* field studies suggest that changes in frequencies of inversions and single-gene alleles can occur rapidly (25) as well as seasonally (26, 27). On the basis of distribution records, it is thought that *D. melanogaster* first entered Australia from the north about 100 years ago (28). Despite the relative recency of this introduction and molecular data indicating a high rate of gene flow (29), clinal variation is well established for several traits and several genetic polymorphisms along the Australian eastern coast (3, 30). Therefore, the genetic shift documented here is unlikely to be the tail end of adaptation to a temporally stable environmental transect by an introduced species; instead, it may be part of a rapid genetic response of a species to climate change, as has recently been suggested for an adaptive trait in another insect species (31).

The *Adh* enzyme polymorphism in *D. melanogaster* and the common cosmopolitan inversions in this species are valuable examples of clinal variation along latitudinal gradients, resulting in genetic constitutions appropriate

to local climatic conditions. The shift in polymorphisms with climatic change indicates how *Adh* [and potentially *In(3R)Payne*] can be linked to the dynamics of adaptive processes. We have shown that adaptive polymorphisms could provide monitoring tools for detecting the impact of climate change on populations.

References and Notes

1. J. A. Endler, *Natural Selection in the Wild* (Princeton Univ. Press, Princeton, NJ, 1986).
2. J. A. Coyne, E. Beecham, *Genetics* **117**, 727 (1987).
3. A. A. Hoffmann, A. R. Anderson, R. Hallas, *Ecol. Lett.* **5**, 614 (2002).
4. J. R. David, C. Bocquet, *Experientia* **31**, 164 (1975).
5. J. G. Oakeshott et al., *Evol. Int. J. Org. Evol.* **36**, 86 (1982).
6. J. Van't Land, W. F. Van Putten, H. Villarreal, A. Kamping, W. Van Delden, *Evol. Int. J. Org. Evol.* **54**, 201 (2000).
7. W. R. Knibb, *Genetica* **58**, 213 (1982).
8. P. R. Anderson, J. G. Oakeshott, *Nature* **308**, 729 (1984).
9. A. Berry, M. Kreitman, *Genetics* **134**, 869 (1993).
10. C. L. Vigue, F. M. Johnson, *Biochem. Genet.* **9**, 213 (1973).
11. J. A. McKenzie, S. W. McKechnie, in *Genetic Studies of Drosophila Populations*, J. B. Gibson, J. G. Oakeshott, Eds. (ANU Press, Canberra, 1983), pp. 201–215.
12. S. N. Alahiitis, *Genetica* **59**, 81 (1982).
13. W. van Delden, A. C. Boerema, A. Kamping, *Genetics* **90**, 161 (1978).
14. See supporting data on Science Online.
15. P. R. Anderson, W. R. Knibb, J. G. Oakeshott, *Genetica* **75**, 81 (1987).
16. A. Kamping, W. Van Delden, *J. Evol. Biol.* **12**, 809 (1999).
17. P. Andolfatto, J. D. Wall, M. Kreitman, *Genetics* **153**, 1297 (1999).
18. J. G. Oakeshott, S. W. McKechnie, G. K. Chambers, *Genetica* **63**, 21 (1984).
19. A. R. Anderson, A. A. Hoffmann, S. W. McKechnie, P. A. Urmina, A. R. Weeks, *Mol. Ecol.*, in press.
20. J. W. Zillman, in *Year Book Australia* (Australian Bureau of Statistics, Canberra, 2003).
21. J. T. Houghton et al., Eds., *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, New York, 2001).
22. Australian Government, Bureau of Meteorology (www.bom.gov.au).
23. W. J. Etges, M. Levitan, *Biol. J. Linn. Soc.* **81**, 395 (2004).
24. F. Rodriguez-Trelles, M. A. Rodriguez, *Evol. Ecol.* **12**, 829 (1998).
25. W. van Delden, in *Population Biology and Evolution*, K. Wohrmann, V. Loeschcke, Eds. (Springer-Verlag, Heidelberg, 1984), pp. 127–142.
26. T. Dobzhansky, *Genetics* **28**, 162 (1943).
27. K. Nielsen, A. A. Hoffmann, S. W. McKechnie, *Genet. Sel. Evol.* **17**, 41 (1985).
28. I. R. Bock, P. A. Parsons, in *Genetics and Biology of Drosophila*, M. Ashburner, H. L. Carsons, J. N. Thompson, Eds. (Academic Press, London, 1981), pp. 299–308.
29. W. J. Kennington, J. Gockel, L. Partridge, *Genetics* **165**, 667 (2003).
30. A. R. Weeks, S. W. McKechnie, A. A. Hoffmann, *Ecol. Lett.* **5**, 756 (2002).
31. W. E. Bradshaw, C. M. Holzapfel, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14509 (2001).
32. We thank A. Anderson, K. Viduka, J. Griffiths, and V. Kellerman for technical assistance and fly collecting. Supported by the Australian Research Council via their Special Research Centre Program. This work was carried out while A.R.W. and M.R.K. were the recipients of Australian Research Council postdoctoral fellowships.

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5722/691/DC1

Materials and Methods
References

7 January 2005; accepted 9 February 2005
10.1126/science.1109523

PERIOD1-Associated Proteins Modulate the Negative Limb of the Mammalian Circadian Oscillator

Steven A. Brown,^{1*} Juergen Ripperger,¹ Sebastian Kadener,²
Fabienne Fleury-Olela,¹ Francis Vilbois,³ Michael Rosbash,²
Ueli Schibler^{1*}

The clock proteins PERIOD1 (PER1) and PERIOD2 (PER2) play essential roles in a negative transcriptional feedback loop that generates circadian rhythms in mammalian cells. We identified two PER1-associated factors, NONO and WDR5, that modulate PER activity. The reduction of NONO expression by RNA interference (RNAi) attenuated circadian rhythms in mammalian cells, and fruit flies carrying a hypomorphic allele were nearly arrhythmic. WDR5, a subunit of histone methyltransferase complexes, augmented PER-mediated transcriptional repression, and its reduction by RNAi diminished circadian histone methylations at the promoter of a clock gene.

About 10% of all mammalian transcripts show daily oscillations of abundance (1). These rhythmic fluctuations are governed by a molecular circadian clock, whose function relies on two interconnected feedback loops of transcription. In the major negative feedback loop, transcription of the *Period* (*Per1* and *Per2*) and *Cryptochrome* (*Cry1* and *Cry2*) genes, and of *Rev-Erba*, an orphan nuclear receptor gene, is activated by the transcription factors CLOCK and BMAL1 and repressed by the PERIOD

(PER) and CRYPTOCHROME (CRY) proteins themselves (2). Although PERs interact with CRYs (3), the mechanism by which they repress CLOCK:BMAL1-mediated clock gene transcription remains poorly understood.

To elucidate the sizes of the PER1 and PER2 complexes involved in this process, we fractionated nuclear extracts of mouse livers by gel filtration chromatography. During the night, the two proteins formed similarly large complexes (>1 MD) (Fig. 1A), whose abundance

and size distribution changed during the day (fig. S1). Thus, PERs associate with other proteins that may play roles in the function of the circadian oscillator. To identify them, we generated a Rat-1 fibroblast cell line that expresses a PER1 protein containing a 6xHis tag and a V5 epitope tag at its C terminus. The expression of a His-V5-PER1 protein was three times higher than that of endogenous PER1 in unsynchronized cells (Fig. 1B, lanes 1 and 2) and did not interfere with the circadian transcription of other clock or clock-controlled genes (4).

Nuclear extracts from His-V5-PER1 cells were harvested 3 hours after the induction of circadian rhythms by serum treatment. Because clock-controlled genes such as *Dbp* and *Rev-erba* are repressed during this time period (5), PER-mediated transcriptional repression would likely be operative as well. After chromatography of this extract was performed on nickel

¹Department of Molecular Biology and National Centres of Competence in Research (NCCR) Frontiers in Genetics, Sciences III, University of Geneva, 30 Quai Ernest Ansermet, CH-1211 Geneva-4, Switzerland.

²Department of Biology, Howard Hughes Medical Institute, Brandeis University, Waltham, MA 02454, USA. ³Serono Pharmaceutical Research Institute, 14 Chemin des Aulx, 1228 Plan-les-Ouates, Geneva, Switzerland.

*To whom correspondence should be addressed. E-mail: steven.brown@molbio.unige.ch (S.A.B.); ueli.schibler@molbio.unige.ch (U.S.)

chelate resin, superose 6, and V5 antibody–Sephacrose (fig. S2A), the eluted proteins were size-fractionated by gel electrophoresis and stained with colloidal Coomassie blue (Fig. 1B, lane 3). Individual bands were excised, and the peptides they contained were identified by tandem mass spectrometry. In addition to peptides from CRY1 and CRY2, which are already known to be PER1-interacting proteins (3), peptides from two other factors, NONO and WDR5, were identified. NONO has been characterized previously as an RNA- and DNA-binding protein that could be involved in splicing, transcriptional repression, and RNA export (6); and its *Drosophila* homolog, NonA, has been implicated in visual acuity and

courtship behavior (7). WDR5 is a member of a histone methyltransferase complex (8) and has been implicated in cell differentiation processes (9). RNA and proteins for both factors are expressed at constant levels throughout the day in a variety of tissues, including the suprachiasmatic nucleus, which is the site of the central circadian clock in the mammalian hypothalamus (fig. S3). However, the apparent sizes of the NONO protein complexes in liver nuclei varied during the day, and a size difference of these complexes was also observed between wild-type mice and mice lacking functional PER proteins (Fig. 1C).

To verify the association of WDR5 and NONO with PER1, gel filtration chromato-

graphy was performed on cellular extracts. A portion of the NONO protein coeluted with PER1, and complete immunodepletion of His-V5-PER1 before fractionation resulted in a decrease in apparent size for some NONO proteins (fig. S2B, fractions 56 to 58). Because the majority of the WDR5 protein was present in complexes of sizes close to those that contain PER1, and because we estimate WDR5 to be 10 times more abundant than PER1 in the cell, it was impossible to estimate the fraction of WDR5 that co-migrates with PER1 by using this assay (4). Nevertheless, epitope-tagged PER1 or PER2 coimmunoprecipitated with WDR5 and NONO (Fig. 1D), and both WDR5 and NONO were immunologically de-

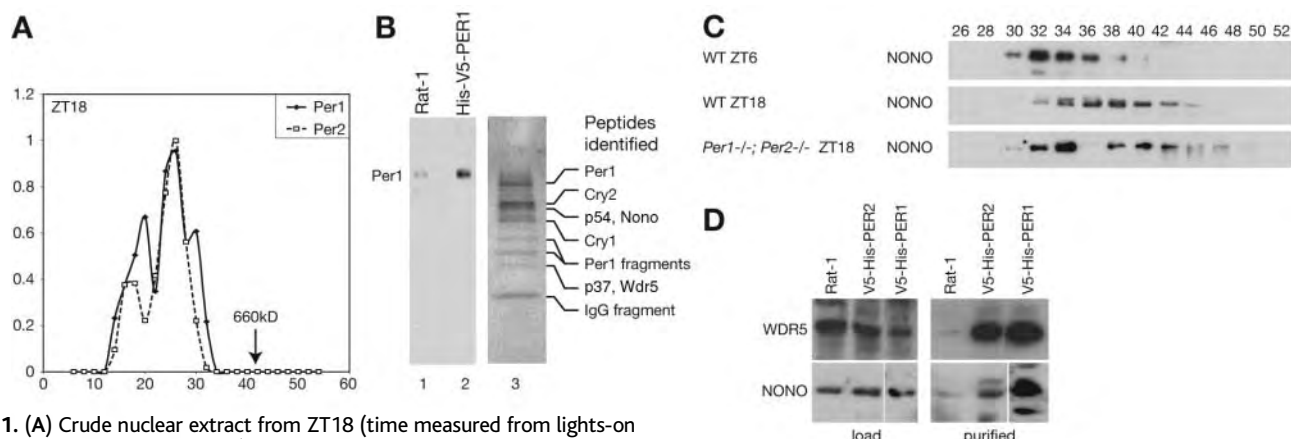


Fig. 1. (A) Crude nuclear extract from ZT18 (time measured from lights-on in a 12-hour light-dark cycle) was fractionated by gel filtration chromatography (Biogel A1.5m, Bio-Rad, Hercules, CA), and fractions were analyzed by Western blot probed with anti-PER1 and anti-PER2. x axis, fraction number (smaller numbers are larger complexes); y axis, signal intensity in arbitrary units. The elution peak of thyroglobulin (660 kD) is shown for reference (see also fig. S1). (B) Lanes 1 and 2 show whole-cell extracts from RAT1 cells and His-V5-PER1 cells that were analyzed on a Western blot probed with anti-PER1. Lane 3 shows colloidal Coomassie-stained gel of purified material; peptides identified by mass spectrometry

are shown at right. (C) Nuclear extract from wild-type mice killed at ZT6 (top) or ZT18 (middle), or from *Per1*^{-/-}*Per2*^{-/-} double-knockout mice (bottom) was purified by nickel-nitriloacetic acid (NiNTA) chromatography and then analyzed on a Western blot probed with anti-NONO. (D) Extracts from serum-treated RAT1 cells, V5-His-PER1 cells, or V5-His-PER2 cells were purified by NiNTA chromatography and then analyzed by Western blot with anti-WDR5 (top) or with anti-NONO (bottom). (Left) One-fiftieth of the crude extract used; (right) purified.

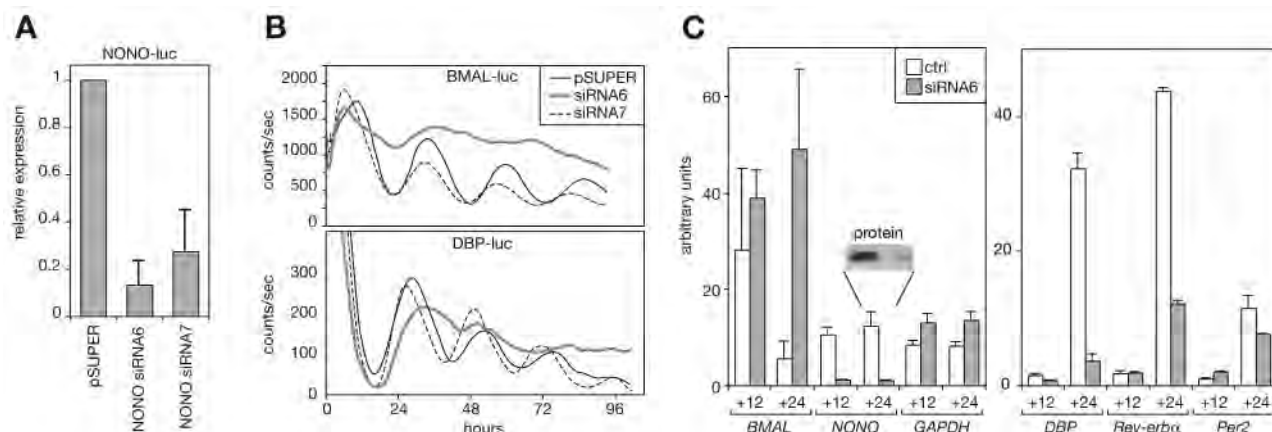


Fig. 2. (A) 3T3 cells were transfected with *NONO*-luciferase fusions, as well as the pSUPER empty vector or the *NONO*-specific siRNA hairpin expression vectors pSUPER-*NONO*-siRNA6 or pSUPER-*NONO*-siRNA7. *NONO*-luciferase levels were measured by luminometry of whole-cell extracts. (B) 3T3 cells were transfected with *NONO*-luciferase fusions, as well as with the pSUPER blank siRNA hairpin expression vector or the *NONO*-specific hairpin vectors pSUPER-*NONO*-siRNA6 or pSUPER-*NONO*-siRNA7. After synchronization of circadian rhythms with dexamethasone, *BMAL11*-luciferase (top) and *DBP*-luciferase (bottom)

expression levels were measured by continuous live-cell luminometry. (C) Cells cotransfected with pSUPER or pSUPER-*NONO*-siRNA6 and the pCS2 YFP-Venus expression vector were treated with dexamethasone and then FACS-sorted for the presence of YFP 12 or 24 hours later. RNA from YFP-positive cells was analyzed by real-time quantitative polymerase chain reaction (PCR) by using indicated probes and normalized as described in (20). In addition, *NONO* protein in sorted whole-cell extract from duplicate plates was measured by Western blot (inset). GAPDH, glyceraldehyde phosphate dehydrogenase; ctrl, control.

tectable after chromatography as per the initial purification (fig. S2C).

To determine whether WDR5 and NONO control circadian gene expression, RNA interference (RNAi) experiments were conducted to reduce target protein levels. For NONO, 12 short RNAi hairpin sequences (10) were screened for their ability to reduce luciferase expression in mouse 3T3 fibroblasts that were transiently transfected with a plasmid that expresses a luciferase-NONO hybrid transcript. Two of these hairpins, which elicited 4- or 10-fold reductions in luciferase levels (Fig. 2A), were expressed in 3T3 cells with the circadian reporter plasmids pBmal1-luc or pDbp-luc that express luciferase in opposite circadian phases. The circadian rhythms of transfected cells were synchronized (11), and circadian cycles of *Bmal-luc* and *Dbp-luc* expression were measured by real-time luminometry (12). The RNAi hairpin vector that reduced *NONO-luc* transcription fourfold shortened the circadian period of transcription by approximately 2 hours.

The other hairpin, which reduced *NONO-luc* transcription 10-fold, attenuated circadian transcription of both reporter genes by at least 20 times and increased the expression of the *Bmal-luc* fusion construct (Fig. 2B).

To measure the effect of NONO depletion on endogenous gene expression, 3T3 cells were cotransfected with a *NONO*-small interfering RNA (siRNA) vector and a yellow fluorescent protein (YFP) vector, synchronized, and then sorted with a fluorescence-activated cell sorter (FACS) to purify YFP-expressing cells at two different times of day. Endogenous *NONO* transcript and protein levels were reduced in these cells. *Bmal1* transcript levels exhibited no temporal variation and were constantly high. However, transcript levels of *Rev-erba* and *Dbp* were constantly low (Fig. 2C), suggesting that the wild-type *NONO* protein does not aid in the repression of these genes but rather antagonizes it.

NonA, the *Drosophila* homolog of *NONO*, plays an important physiological role, because

flies harboring a P element-mediated deletion of this gene are extremely sick (13). However, because flies homozygous for the strongly hypomorphic *nonA* allele *P14* are viable and fertile (7), we were able to examine whether *NonA* also plays a role in the generation of circadian rhythms in *Drosophila*. *nonA^{P14/P14}* flies were behaviorally nearly arrhythmic and hyperactive (Fig. 3A and fig. S4). In addition, the rhythmicity of mRNA from the clock gene *timeless* was greatly attenuated, indicating that the *nonA^{P14}* mutation interferes with the basic function of the circadian oscillator (Fig. 3C).

The WDR5 protein has been identified as a member of a histone methyltransferase complex (8). Consistent with this function, histone methyltransferase activity could be precipitated from His-V5-PER1 fibroblast extracts with antibodies to V5 (anti-V5), anti-PER2, and anti-WDR5. It could also be enriched from evening but not morning nuclear extracts of mouse livers with anti-PER2 (Fig. 4A). Because PER-regulated gene expression is low at night, it is likely that this methyltransferase activity has a repressive function. Such a repressive effect would be characteristic of histone H3 lysine 9 (H3K9) methylation, among other types (14). Nevertheless, WDR5 has been associated with a histone H3K4 methyltransferase containing a SET1 domain (8), which is mostly thought to be involved in the activation of transcription (15). Hence, both histone modifications were examined at a PER1-regulated circadian promoter in the presence and absence of WDR5.

A lentivirally mediated system (16) was used to generate a cell line that represses endogenous *Wdr5* RNA and protein levels about five times in a doxycycline-inducible fashion (fig. S5). Circadian cycling in these cells was synchronized (11), and chromatin was harvested from cells grown in the presence or absence of doxycycline at the time of maximum and minimum *Rev-erba* transcription. Methylation of histone H3K4 and histone H3K9 at the *Rev-erba* promoter was examined by chromatin immunoprecipitation (ChIP). In uninduced cells expressing normal levels of WDR5, circadian methylation was observed in phase with *Rev-erba* transcription at H3K4, and antiphase to it at H3K9. In doxycycline-treated cells, however, both rhythms were nearly abolished (Fig. 4B). Down-regulation of WDR5 expression produced only moderate changes in clock gene expression. PER-regulated genes such as *Per2* and *Rev-erba* were somewhat derepressed initially, and mRNA accumulation was then phase-delayed by about 2 hours (Fig. 4B and fig. S6). Expression of other circadian transcripts such as *Cry1* mRNA, *Clock* mRNA, and *Bmal1* mRNA were unaffected (4). Similarly, when the same cells were analyzed with circadian luciferase reporter genes and real-time luminometry, the period length of cyclic gene ex-

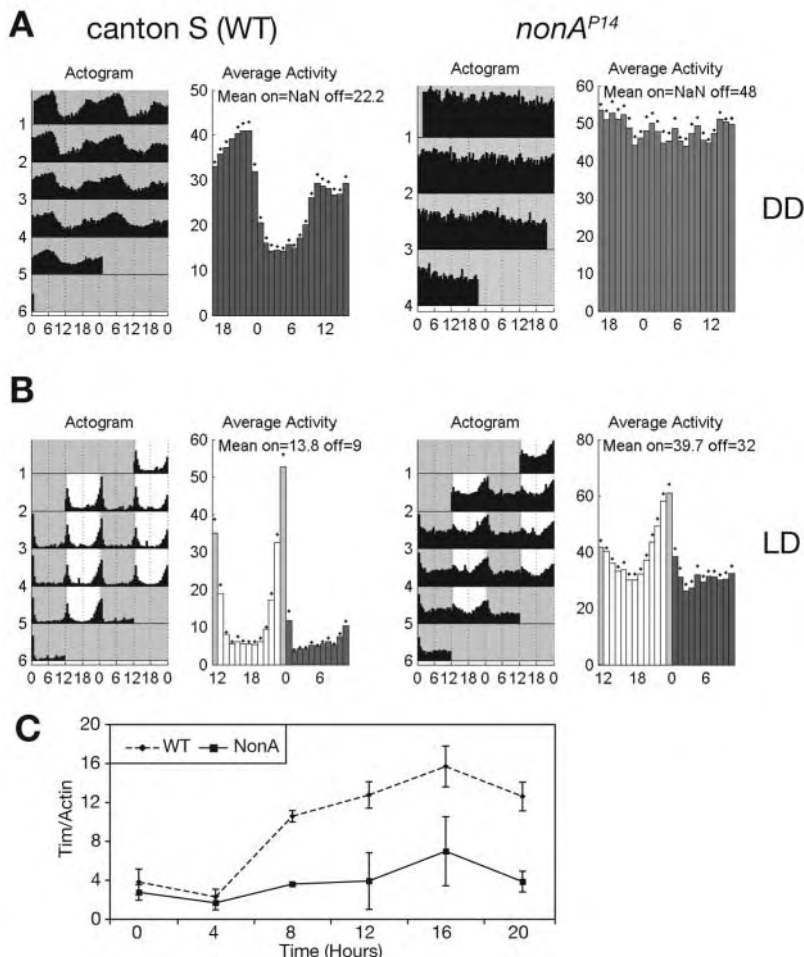
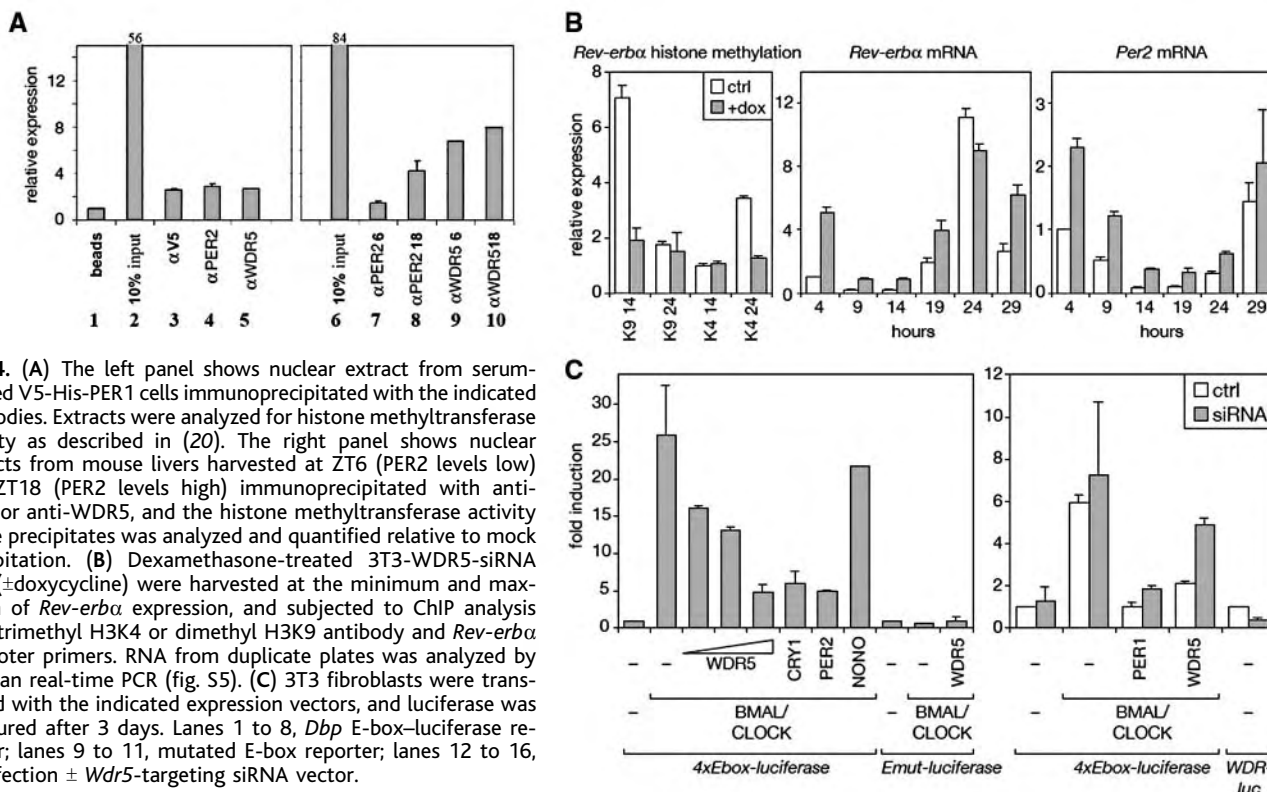


Fig. 3. (A) Locomotor activities of male Canton S or *nonA^{P14/P14}* flies were analyzed by infrared beam-break studies in continuously dark (DD) conditions. Sample data from a single fly are shown at left, and data from 10 flies are plotted at right (fig. S3). (B) Identical experiment to (A) but measured in flies exposed to 12 hours of light and 12 hours of darkness per day (LD). (C) RNA from wild-type and mutant flies was prepared from four separate pools of 10 fly heads harvested at the indicated time of day, and the quantity of *Tim* RNA was determined by Taqman real-time PCR.



pression was unchanged (fig S6), suggesting that these two histone modifications are either compensatory or irrelevant for circadian clock function. Because overall bioluminescence from multiple reporters was also reduced, we conclude that WDR5 reduction has a pleiotropic effect on cellular processes unrelated to circadian clock function. Therefore, although circadian histone methylation was drastically affected, we could not verify a direct role for WDR5 in the circadian clock via our loss-of-function experiments.

PER proteins repress the activity of the circadian activators CLOCK and BMAL in transient transfection assays (17). The transient expression of WDR5 repressed CLOCK:BMAL1-mediated transcription to the same extent as the transient expression of the PER2 or CRY1 proteins, and inhibition of transcription by either PER2 or WDR5 could be partially blocked by cotransfecting a WDR5 siRNA vector. This hairpin reduced expression levels of a *WDR5-luc* fusion transcript threefold in parallel experiments (Fig. 4C). Thus, WDR5 can aid the PER-CRY complex in repressing CLOCK:BMAL1-mediated transcription, although this effect might be indirect. Because NONO was 50 times less efficiently expressed than WDR5, PER2, or CRY1 in these experiments, it was not possible to demonstrate its supposed antagonistic effect by this method.

We have identified two proteins, NONO and WDR5, that can associate with the mammalian PER1 protein. Our data suggest that

NONO probably operates antagonistically to PER proteins in mammalian cells, and that it is essential to normal circadian rhythmicity in mammals and in *Drosophila*. Mutations in the *Drosophila* homologs of PER1 and NONO were previously coincided in a screen for courtship behavior defects (18). Overexpression experiments suggest that WDR5 assists PER function in mammalian cells. Its reduction affected two different antiphasic circadian histone modifications, H3K4 and H3K9, which are thought to have opposite effects (14, 15), but WDR5 had only moderate consequences for circadian clock function. Hence, further studies of WDR5 and of histone posttranslational modifications at circadian promoters will be essential for understanding its contribution to clock function. We suggest that the antagonistic activities of these proteins might help to render the circadian clock more resilient to noise—caused by changes in temperature and nutrients, or by stochastic fluctuations in transcription rates—that could affect the concentrations of critical clock components (19).

References and Notes

1. P. L. Lowrey, J. S. Takahashi, *Annu. Rev. Genomics Hum. Genet.* **5**, 407 (2004).
2. U. Albrecht, G. Eichele, *Curr. Opin. Genet. Dev.* **13**, 271 (2003).
3. C. Lee, J. P. Etchegaray, F. R. Cagampang, A. S. Loudon, S. M. Reppert, *Cell* **107**, 855 (2001).
4. S. A. Brown *et al.*, data not shown.
5. A. Balsalobre, F. Damiola, U. Schibler, *Cell* **93**, 929 (1998).
6. Y. Shav-Tal, D. Zipori, *FEBS Lett.* **531**, 109 (2002).

7. K. G. Rendahl, K. R. Jones, S. J. Kulkarni, S. H. Bagully, J. C. Hall, *J. Neurosci.* **12**, 390 (1992).
8. J. Wysocka, M. P. Myers, C. D. Laherty, R. N. Eisenman, W. Herr, *Genes Dev.* **17**, 896 (2003).
9. F. Gori, P. Divieti, M. B. Demay, *J. Biol. Chem.* **276**, 46515 (2001).
10. P. J. Paddison, A. A. Caudy, E. Bernstein, G. J. Hannon, D. S. Conklin, *Genes Dev.* **16**, 948 (2002).
11. A. Balsalobre *et al.*, *Science* **289**, 2344 (2000).
12. S. Yamazaki *et al.*, *Science* **288**, 682 (2000).
13. R. Stanewsky, K. G. Rendahl, M. Dill, H. Saumweber, *Genetics* **135**, 419 (1993).
14. M. Lachner, T. Jenuwein, *Curr. Opin. Cell Biol.* **14**, 286 (2002).
15. H. Santos-Rosa *et al.*, *Nature* **419**, 407 (2002).
16. M. Wiznerowicz, D. Trono, *J. Virol.* **77**, 8957 (2003).
17. X. Jin *et al.*, *Cell* **96**, 57 (1999).
18. C. P. Kyriacou, J. C. Hall, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6729 (1980).
19. P. S. Swain, M. B. Elowitz, E. D. Siggia, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12795 (2002).
20. Materials and methods are available as supporting material on Science Online.
21. Molecular interaction data have been deposited in the Biomolecular Interaction Network Database (BIND) with accession codes 258456 and 258459. We thank D. Trono and M. Wiznerowicz for materials and counsel about lentiviral transduction and RNAi, W. Herr for advice and antibodies to WDR5, and U. Albrecht for *Per1*^{-/-} and *Per2*^{-/-} mice. Research in the Schibler lab was supported by grants from the Swiss National Science Foundation, the NCCR Frontiers in Genetics research program, and the Jeantet Foundation. The work of S.K. and M.R. was supported by the Howard Hughes Medical Institute and that of F.V. by the Serono Research Foundation.

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5722/693/DC1

Materials and Methods

Figs. S1 to S6

References

10 November 2004; accepted 17 February 2005
10.1126/science.1107373

Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance

Roger Guimerà,^{1*} Brian Uzzi,^{2*} Jarrett Spiro,³
Luís A. Nunes Amaral^{1†}

Agents in creative enterprises are embedded in networks that inspire, support, and evaluate their work. Here, we investigate how the mechanisms by which creative teams self-assemble determine the structure of these collaboration networks. We propose a model for the self-assembly of creative teams that has its basis in three parameters: team size, the fraction of newcomers in new productions, and the tendency of incumbents to repeat previous collaborations. The model suggests that the emergence of a large connected community of practitioners can be described as a phase transition. We find that team assembly mechanisms determine both the structure of the collaboration network and team performance for teams derived from both artistic and scientific fields.

Teams are assembled because of the need to incorporate individuals with different ideas, skills, and resources. Creativity is spurred when proven innovations in one domain are introduced into a new domain, solving old problems and inspiring fresh thinking (1–4). However, research shows that the right balance of diversity on a team is elusive. Although diversity may potentially spur creativity, it typically promotes conflict and miscommunication (5–7). It also runs counter to the security most individuals experience in working and sharing ideas with past collaborators (8). Successful teams evolve toward a size that is large enough to enable specialization and effective division of labor among teammates but small enough to avoid overwhelming costs of group coordination (9). Here, we investigate empirically and theoretically the mechanisms by which teams of creative agents are assembled. We also investigate how these microscopic team assembly mech-

anisms determine both the macroscopic structure of a creative field and the success of certain teams in using the resources and knowledge available in the field. We develop a model for the assembly of teams of creative agents in which the selection of the members of a team is controlled by three parameters: (i) the number, m , of team members; (ii) the probability, p , of selecting incumbents, that is, agents already belonging to the network; and (iii) the propensity, q , of incumbents to select past collaborators. The model predicts the existence of two phases that are determined by the values of m , p , and q . In one phase, there is a large cluster connecting a substantial fraction of the agents, whereas in the other phase the agents form a large number of isolated clusters.

We analyzed data from both artistic and scientific fields where collaboration needs have experienced pressures such as differentiation and specialization, internationalization, and commercialization (4, 10, 11): (i) the Broadway musical industry (BMI) and (ii) the scientific disciplines of social psychology, economics, ecology, and astronomy (Table 1). For the BMI, we considered all 2258 productions in the period from 1877 to 1990 (12, 13). Productions are defined as musical shows that were performed at least once in Broadway. The team members com-

prise individuals responsible for composing the music, writing the libretto and the lyrics, designing the choreography, directing, and producing the show, but not the actors that performed in it. For each of the scientific disciplines, we considered all collaborations that resulted in publications in recognized journals within the fields studied (14): seven social psychology journals, nine economics journals, 10 ecology journals, and six astronomy journals (Table 2). Collaboration networks (15–19) were then built for each of the journals independently and for the whole discipline by merging the data from the journals within a discipline (Materials and Methods).

The evolution of team sizes in the BMI bears out the expectation that team size and composition depend on the intricacy of the creative task. In the period from 1877 to 1929, when the form of the Broadway musical show was still being worked out through trial and error (12), there was a steady increase in the number of artists per production, from an average of two to an average of seven (Fig. 1A). This increase in size suggests that teams evolved to manage the complexity of the new artistic form. By the late 1920s, the Broadway musical reached the form we know today, as did team composition (4). Since then, the typical set of artists creating a Broadway musical have been choreographer, composer, director, librettist, lyricist, and producer. For the following 55 years, a period that includes the Great Depression, World War II, and the postwar boom, the average size of teams remained around seven (20).

We find similar scenarios for the evolution of team size in scientific collaborations. The four fields experienced an increase in team size with time (Fig. 1, B to E). The increase has been roughly linear in social psychology and economics and faster than linear in ecology and astronomy. For social psychology, team size growth rate was greater for high-impact compared with low-impact journals, suggesting that team size not only depends on the intricacy of the enterprise but also that successful teams might adapt faster to external pressures.

The analysis of team size cannot capture the fact that teams are embedded in a larger network (3). This complex network (21–26), which is the result of past collaborations and

¹Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA. ²Kellogg School of Management and Department of Sociology, Northwestern University, Evanston, IL 60208, USA. ³Graduate School of Business, Stanford University, Stanford, CA 94305, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: amaral@northwestern.edu

Table 1. Global network properties of the fields studied. The sources for the BMI are (12) and (13). The data analyzed excludes revivals and focus on the steady-state period from 1940 to 1985. The data for scientific publications was obtained from the Web of Science. We selected recognized journals in each of the different scientific fields (Table 2). For each field, we show the total number of productions

and agents in all the periods considered, the values of p and q estimated with the model from the data, the f_R , the size, N , of the network in the last year of the period considered, the value, N_{mod} , predicted by the model, the fraction, S , of agents that belong to the largest cluster, and the value, S_{mod} , predicted by the model. S takes values between 0 and 1 and does not depend on the size of the network (31).

Field	Period	Productions	Agents	p	q	f_R	N	N_{mod}	S	S_{mod}
BMI	1877–1990	2258	4113	0.52	0.77	0.16	428	420	0.70	0.80
Social psychology	1955–2004	16,526	23,029	0.56	0.78	0.22	11,412	14,408	0.68	0.67
Economics	1955–2004	14,870	23,236	0.57	0.73	0.22	9527	11,172	0.54	0.50
Ecology	1955–2004	26,888	38,609	0.59	0.76	0.23	23,166	26,498	0.75	0.84
Astronomy	1955–2004	30,552	30,192	0.76	0.82	0.39	18,021	22,794	0.92	0.98

the medium in which future collaborations will develop, acts as a storehouse for the pool of knowledge created within the field. The way the members of a team are embedded in the larger network affects the manner in which they access the knowledge in the field. Therefore, teams formed by individuals with

large but disparate sets of collaborators are more likely to draw from a more diverse reservoir of knowledge. At the same time and for the same reasons, the way teams are organized into a larger network affects the likelihood of breakthroughs occurring in a given field.

The agents composing a team may be classified according to their experience. Some agents are newcomers, that is, rookies, with little experience and unseasoned skills. Other agents are incumbents. They are established persons with a track record, a reputation, and identifiable talents. The differentiation of agents into newcomers and incumbents results in four possible types of links within a team: (i) newcomer-newcomer, (ii) newcomer-incumbent, (iii) incumbent-incumbent, and (iv) repeat incumbent-incumbent. The distribution of different types of links reflects the team's underlying diversity. For example, if teams have a preponderance of repeat incumbent-incumbent links, it is less likely that they will have innovative ideas because their shared experiences tend to homogenize their pool of knowledge. In contrast, teams with a variety of types of links are likely to have more diverse perspectives to draw from and therefore to contribute more innovative solutions.

Because quantifying the emergence and the effects of team diversity (2, 9, 27–29) is more difficult than measuring team size, we consider next a model for the assembly of teams. In our model, we assemble N teams in temporal sequence. The assembly of each team is controlled by three parameters: m , p , and q . The first parameter, m , is the number of agents in a team. In our investigations of the model, we considered three situations: (i) keep m constant, (ii) draw m from a distribution, or (iii) use a sequence of m values obtained from the data. For the theoretical analysis of the model, we kept m constant, whereas comparison with an empirical data set was done with the use of the sequence of $m(t)$ values in the corresponding data set.

The second parameter, p , is the probability of a team member being an incumbent. Higher values of p indicate fewer opportunities for newcomers to enter a field. The third parameter, q , represents the inclination for incumbents to collaborate with prior collaborators rather than initiate a new collaboration with an incumbent they have not worked with in the past.

We start at time zero with an endless pool of newcomers. Newcomers become incumbents the first time step after being selected for a team. Each time step t , we assemble a new team and add it to the network (Fig. 2). We select sequentially $m(t)$ different agents. Each agent in a team has a probability, p , of being drawn from the pool of incumbents and a probability, $1 - p$, of being drawn from the pool of newcomers. If the agent is drawn from the incumbents' pool and there is already another incumbent in the team, then (i) with probability q the new agent is randomly selected from among the set of collaborators of a randomly selected incumbent already in the team; (ii) otherwise, he or she is selected at random among all incumbents in the network.

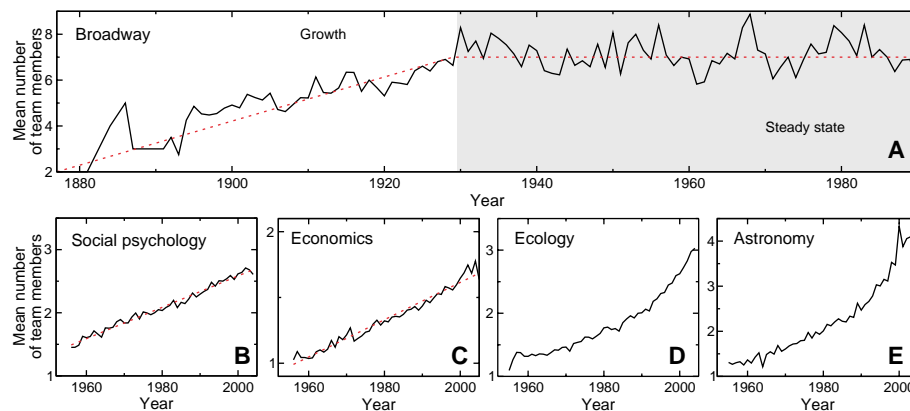


Fig. 1. Time evolution of the typical number of team members in (A) the BMI and scientific collaborations in the disciplines of (B) social psychology, (C) economics, (D) ecology, and (E) astronomy.

Table 2. Journal-specific network structure. We present the same information as in Table 1 for each of the journals studied. We ranked journals within each field according to their impact factor (IF). For some low-impact journals, the f_R is too high to be reproducible with the model. In those cases, which we represent by $q > 1$, simulations of the model are done with $q = 1$. The model still reproduces the empirical results quite well for these cases.

Journal	IF	Period	Agents	p	q	f_R	S	S_{mod}
<i>Social psychology</i>								
<i>J. Pers. Soc. Psychol.</i>	3.862	1965–2003	9112	0.56	0.74	0.20	0.75	0.79
<i>J. Exp. Soc. Psychol.</i>	2.131	1965–2004	2133	0.40	0.76	0.11	0.44	0.07
<i>Pers. Soc. Psychol. B</i>	1.839	1976–2004	4339	0.45	0.74	0.14	0.54	0.47
<i>Eur. J. Soc. Psychol.</i>	1.060	1971–2004	1790	0.41	0.93	0.15	0.44	0.08
<i>J. Appl. Soc. Psychol.</i>	0.523	1971–2004	4602	0.33	1.00	0.10	0.06	0.02
<i>J. Soc. Psychol.</i>	0.291	1956–2004	6294	0.32	>1	0.12	0.05	0.01
<i>Soc. Behav. Personal.</i>	0.227	1973–2004	1981	0.26	>1	0.08	0.03	0.01
<i>Economics</i>								
<i>Q. J. Econ.</i>	4.756	1956–2004	2320	0.37	0.58	0.08	0.26	0.05
<i>Econometrica</i>	2.215	1965–2004	3351	0.45	0.67	0.13	0.26	0.05
<i>J. Polit. Econ.</i>	2.196	1956–2004	3464	0.30	0.88	0.07	0.06	0.01
<i>Am. Econ. Rev.</i>	1.938	1956–2004	6807	0.42	0.84	0.15	0.27	0.02
<i>Econ. J.</i>	1.295	1956–2004	4528	0.31	0.99	0.09	0.08	0.01
<i>Eur. Econ. Rev.</i>	1.021	1969–2004	2585	0.35	0.85	0.10	0.15	0.02
<i>J. Econ. Theory</i>	0.833	1969–2004	2062	0.28	>1	0.08	0.51	0.03
<i>Econ. Lett.</i>	0.337	1978–2004	5129	0.31	0.98	0.10	0.01	0.01
<i>Appl. Econ.</i>	0.200	1969–2004	4488	0.26	>1	0.08	0.01	0.01
<i>Ecology</i>								
<i>Am. Nat.</i>	4.059	1955–2004	4990	0.44	0.70	0.13	0.49	0.19
<i>Ecology</i>	3.701	1965–2003	8885	0.48	0.71	0.15	0.56	0.65
<i>Oecologia</i>	3.128	1969–2004	10,545	0.44	0.81	0.15	0.51	0.36
<i>Ecol. Appl.</i>	2.852	1991–2004	3417	0.29	0.99	0.08	0.30	0.06
<i>J. Ecol.</i>	2.833	1955–2004	3639	0.43	0.91	0.15	0.40	0.19
<i>Funct. Ecol.</i>	2.351	1989–2004	2873	0.36	>1	0.13	0.05	0.02
<i>Oikos</i>	2.142	1961–2004	6589	0.43	0.84	0.15	0.48	0.11
<i>Biol. Conserv.</i>	2.056	1977–2004	5821	0.27	>1	0.09	0.08	0.01
<i>Ecol. Model.</i>	1.561	1978–2004	5260	0.35	>1	0.13	0.14	0.02
<i>J. Nat. Hist.</i>	0.497	1967–2004	2631	0.36	>1	0.04	0.13	0.01
<i>Astronomy</i>								
<i>Astron. J.</i>	5.647	1965–2003	10,832	0.78	0.86	0.40	0.96	0.99
<i>Publ. Astron. Soc. Pac.</i>	3.529	1955–2004	6769	0.58	0.78	0.22	0.85	0.89
<i>Icarus</i>	2.611	1983–2004	4357	0.72	0.90	0.38	0.89	0.97
<i>Publ. Astron. Soc. Jpn.</i>	2.312	1965–2004	2432	0.77	0.95	0.44	0.95	0.99
<i>Astrophys. Space Sci.</i>	0.522	1968–2004	10,823	0.55	1.00	0.29	0.60	0.05
<i>IAU Symp.</i>	0.237	1984–2004	10,185	0.60	0.75	0.23	0.80	0.92

Lastly, agents that remain inactive for longer than τ time steps are removed from the network. This rule is motivated by the observation that agents do not remain in the network forever: agents age and retire, change careers, and so on. The removal process enables the network to reach a steady state after a transient time. Our results do not depend in the specific value of τ (Materials and Methods).

Through participation in a team, agents become part of a large network (30). This fact prompted us to examine the topology of the network of collaborations among the practitioners of a given field. More specifically, we asked, “Is there a large connected cluster comprising most of the agents or is the network composed of numerous smaller clusters?” A large connected cluster would be supporting evidence for the so-called invisible college, the web of social and professional contacts linking scientists across universities proposed by de Solla Price (31) and Merton (32). A large number of small clusters would be indicative of a field made up of isolated schools of thought. For all five fields considered here, we find that the network contains a large connected cluster.

As is typically done in the study of percolation phase transitions (33), we use the fraction S of agents that belong to the largest cluster of the network to quantify the transition between these two regimes: invisible college or isolated schools. We explore systematically the (p, q) parameter space of the model. We find that the system undergoes a percolation transition (33) at a critical line, $p_c(m, q)$. That is, the system experiences a sharp transition from a multitude of small clusters to a situation in which one large cluster, comprising a substantial fraction S of the individuals, emerges: the so-called giant component (Fig. 3). The transition line $p_c(m, q)$ therefore determines the tipping point for the emergence of the invisible college (34). Our analysis shows that the existence of this transition is independent of the average number of agents (m) in a collaboration, although the precise value of $p_c(m, q)$ does depend on m .

The proximity to the transition line, which depends on the distribution of the different types of links, determines the structure of the largest cluster (Fig. 3A). In the vicinity of the transition, the largest cluster has an almost linear or branched structure (Fig. 3A) ($p = 0.30$). As one moves toward larger p , the largest cluster starts to have more and more loops (Fig. 3A) ($p = 0.35$), and, eventually, it becomes a densely connected network (Fig. 3A) ($p = 0.60$).

Networks with the same fraction, S , of nodes in the largest cluster do not necessarily correspond to networks with identical properties. Each point in the (p, q) parameter space is characterized by both S and the fraction, f_R , of repeat incumbent-incumbent links. For

example, in Fig. 3C, the line $f_R = 0.32$ corresponds to those values of p and q for which 32% of all links in new teams are between repeat collaborators (35). The f_R has a notable impact on the dynamics of the network. When f_R is large, collaborations are firmly established, and therefore the structure of the network changes very slowly. In contrast, low values of f_R correspond to enterprises with high turnover and very fast dynamics. Intermediate values of f_R are related to situations in which collaboration patterns with peers are fluid (Materials and Methods).

For each of the five fields for which we have empirical data, we measure the relative size of the giant component S (Materials and Methods). For all fields considered, S is larger than 50% (Table 1). This result provides quantitative evidence for the existence of an invisible college in all the fields. Intriguingly, the relative sizes of the giant com-

ponent is similar for three of the four fields considered: $S = 0.70$, $S = 0.68$, and $S = 0.75$ for BMI, social psychology, and ecology, respectively. However, for astronomy S was significantly larger (0.92), whereas for economics it was significantly smaller (0.54).

To gain further insight in the structure of collaboration networks, we used our model to estimate the values of p and q for each field. Given the temporal sequence of teams producing the network of collaborations, one can calculate the fraction of incumbents and the fraction of repeat incumbent-incumbent links. These fractions and the model enable us to then estimate the values of p and q that are consistent with the data (36).

We estimated p and q for each field and then simulated the model to predict the key properties of the network of collaborations, including the degree distribution of the network and the fraction S of nodes in the

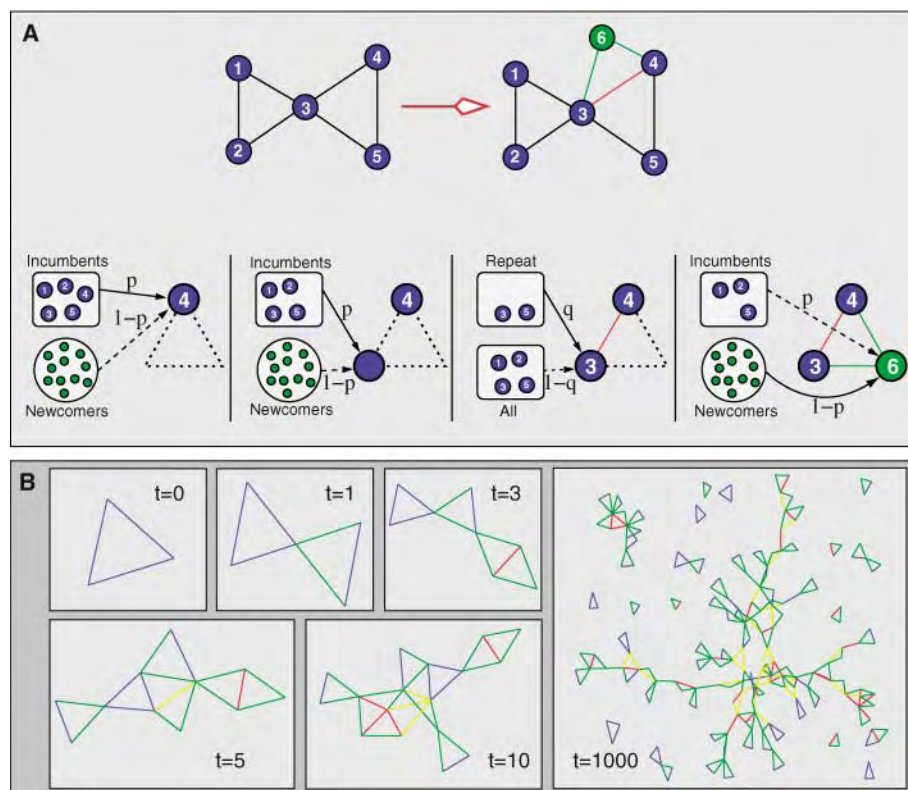


Fig. 2. Modeling the emergence of collaboration networks in creative enterprises. **(A)** Creation of a team with $m = 3$ agents. Consider, at time zero, a collaboration network comprising five agents, all incumbents (blue circles). Along with the incumbents, there is a large pool of newcomers (green circles) available to participate in new teams. Each agent in a team has a probability p of being drawn from the pool of incumbents and a probability $1 - p$ of being drawn from the pool of newcomers. For the second and subsequent agents selected from the incumbents' pool: (i) with probability q , the new agent is randomly selected from among the set of collaborators of a randomly selected incumbent already in the team; (ii) otherwise, he or she is selected at random among all incumbents in the network. For concreteness, let us assume that incumbent 4 is selected as the first agent in the new team (leftmost box). Let us also assume that the second agent is an incumbent, too (center-left box). In this example, the second agent is a past collaborator of agent 4, specifically agent 3 (center-right box). Lastly, the third agent is selected from the pool of newcomers; this agent becomes incumbent 6 (rightmost box). In these boxes and in the following panels and figures, blue lines indicate newcomer-newcomer collaborations, green lines indicate newcomer-incumbent collaborations, yellow lines indicate new incumbent-incumbent collaborations, and red lines indicate repeat collaborations. **(B)** Time evolution of the network of collaborations according to the model for $p = 0.5$, $q = 0.5$, and $m = 3$.

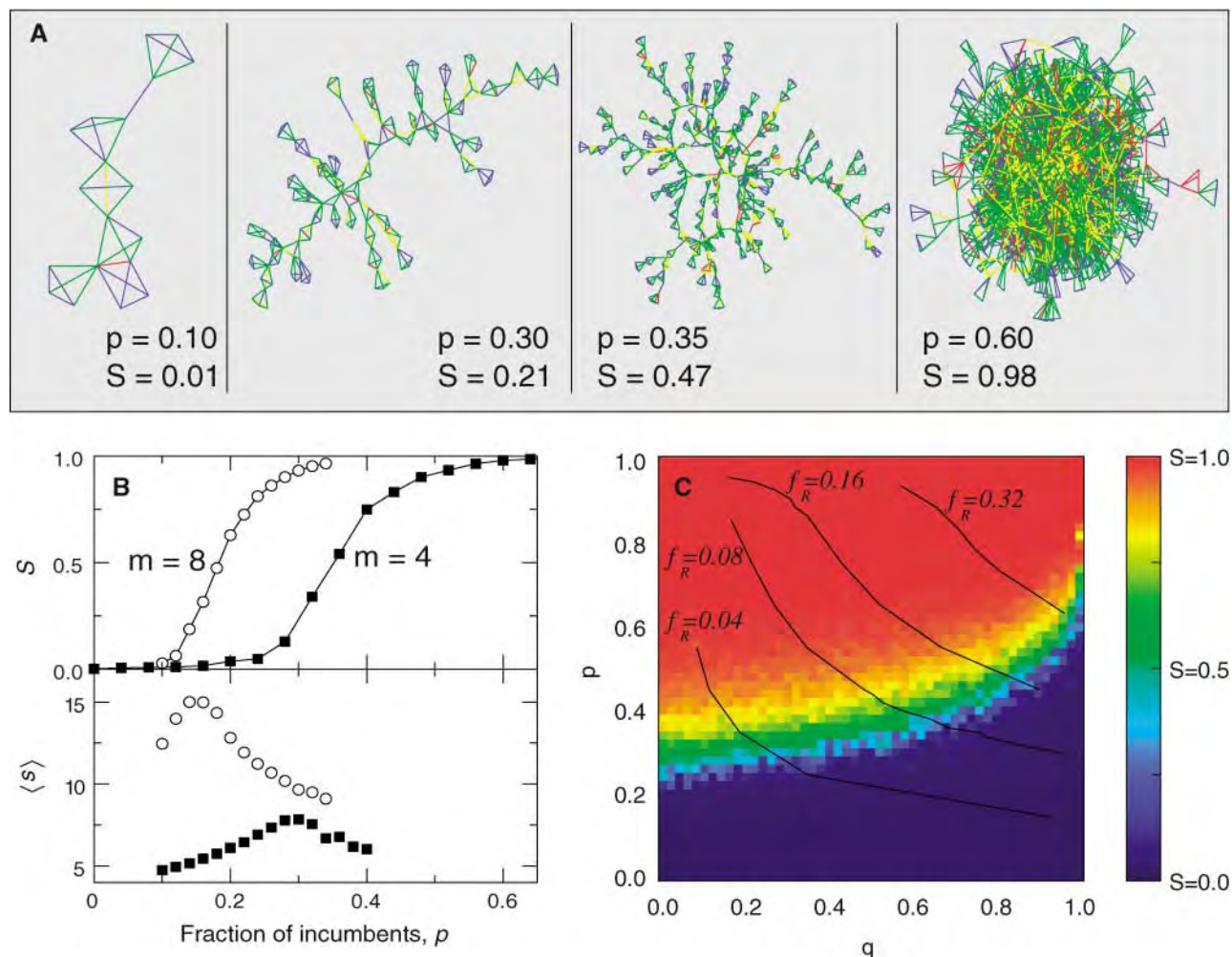


Fig. 3. Predictions of the model. (A) Phase transition in the structure of the collaboration network. We plot only the largest cluster in the network. For small p , the network is formed by numerous small clusters ($p = 0.10$). At the critical point p_c , the tipping point, a large cluster emerges, that is, a cluster that contains a substantial fraction of the agents. In the vicinity of the transition, the largest cluster has an almost linear or branched structure ($p = 0.30$). As p increases, the largest cluster starts to have loops ($p = 0.35$) and eventually becomes a densely connected cluster containing essentially all nodes in the network ($p = 0.60$). We show results for $q = 0.5$ and $m = 4$, where m is the number of agents in a team. (B) The transition described in (A) can be characterized by the fraction S of nodes that belong to the giant component, the order parameter, and the average size $\langle s \rangle$ of the other clusters, the susceptibility (33). The model displays a second-order percolation transition as the fraction p of incumbents increases from 0 to

1. The transition occurs for $p = p_c$, which coincides with the maximum of $\langle s \rangle$. Note that p_c is a decreasing function of m . We show results for $q = 0.5$ and $m = 4$ and $m = 8$. (C) We display graphically the value of S as a function of p and q for $m = 4$. For any value of q , the model displays the percolation transition, and the critical fraction p_c depends on q , defining a percolation line $p_c(m, q)$. The critical line $p_c(m, q)$ is an increasing function of q . Even though the order parameter S is an important parameter to quantify the structure of the network, not all points with the same S , that is, all points represented with the same color, correspond to fields with identical properties. This result is made clear by the lines of equal f_R . The upper-right corner of the (p, q) plane is characterized by f_R close to one, whereas the lower-left corner corresponds to f_R close to zero. As we show in Fig. 4, all fields considered have parameter values above the transition line.

largest cluster. By comparing predictions of the model with the empirical results, we are able to test and validate the model. We first compare the degree distribution of the collaboration networks with the predictions of the model (Fig. 4, A to E) and find that the model predicts the empirical degree distributions remarkably well. In Table 1, we compare the predictions of the model for S with the measured values. The model correctly predicts that an invisible college containing more than 50% of the nodes exists in all cases. Additionally, the values of S predicted by the model are in close agreement with the empirical results.

To investigate how changes of the team assembly mechanism affect the structure of the network, we used the model to generate networks with the same sequence of team sizes as the data but with different values of p and q . We show in Fig. 4, F to J, that four out of the five creative networks we consider are very close to the tipping line at which an invisible college emerges. The exception is astronomy. We also find that, for astronomy, the f_R is significantly larger than for the other fields.

If diversity affects team performance and our model correctly captures how diversity is related to the way teams are assembled, then the parameters p and q must be related to team

performance. To investigate this issue, we considered for the four scientific fields how teams publishing in different journals are assembled. We used each journal's impact factor as a proxy for the typical quality of teams' output. We then studied the different journals separately to quantify the relationship between team assembly mechanisms and performance.

In Fig. 5, we show the values of p , q , and S for the journals in each of the fields as a function of the impact factor of the journal. We found that p was positively correlated with impact factor for economics, ecology, and social psychology, whereas q was negatively correlated with impact factor for the

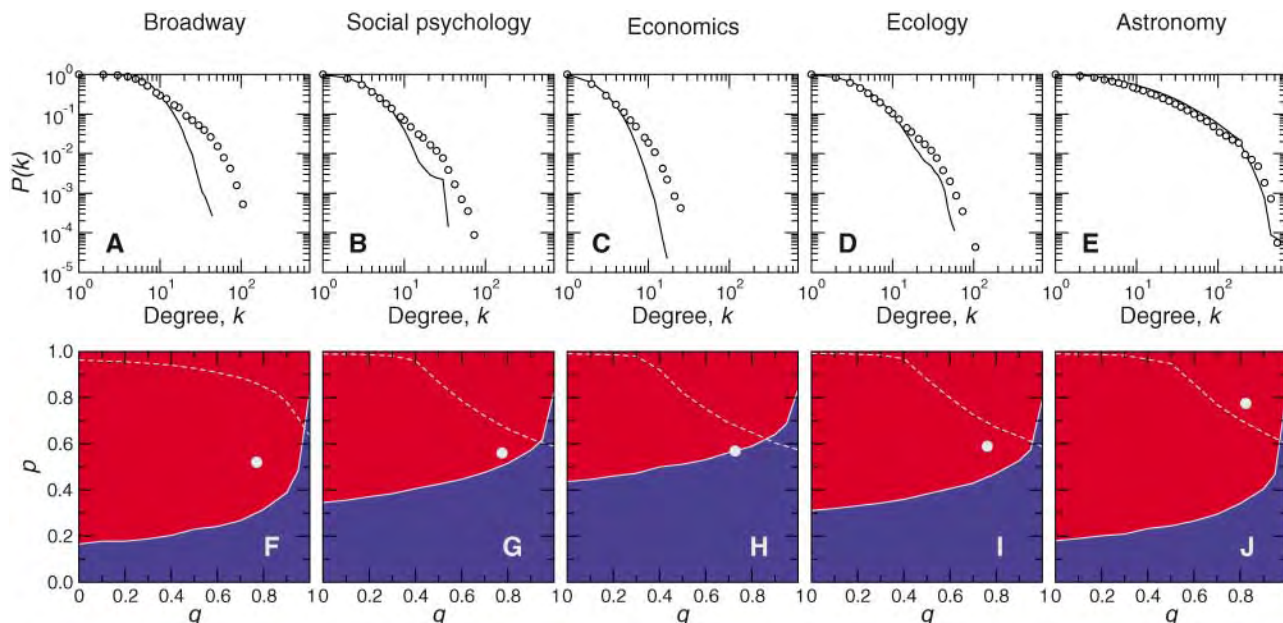


Fig. 4. Network structure of different creative fields. Degree distributions for (A) the BMI, (B) the field of social psychology, (C) the field of economics, (D) the field of ecology, and (E) the field of astronomy. We carried out with the use of the sequence $\{m(t)\}$ of team sizes found in the empirical data and with the values of p and q estimated from the measured fractions of the different types of links. We present the predictions of the model with the lines and the empirical degree distributions with the open circles. For all cases considered, the data falls within the 95% confidence intervals of the predictions of the model. The (p, q) parameter space of the network of collaborators is shown for (F) the BMI, (G) the field of social psychology, (H) the field of economics, (I) the field of ecology, and (J) the field of astronomy. The solid lines

separating the red and the blue regions indicate the values of p and q for which 50% of the nodes belong to the largest cluster, that is, the percolation transition at which a giant component, the invisible college, emerges. The distance from the percolation line predicts the overall structure of the network. For example, the networks in astronomy are well above the tipping line and have a very dense structure (Table 1). In contrast, all other fields are close to the transition and have relatively sparse giant components. Another important characteristic of the network is provided by the value of f_R . To help with the interpretation of the results, we plot with dotted lines the curves for $f_R = 0.32$. For four of the creative networks considered, we find $f_R < 0.25$. For astronomy, we find $f_R = 0.39$.

same fields. The result for p implies that successful teams have a higher fraction of incumbents, who contribute expertise and know-how to the team, whereas the result for q implies that teams that are less diverse typically have lower levels of performance.

The relative size S of the giant component in a journal was also associated with performance for ecology and social psychology. Teams publishing in journals with a high-impact factor typically give rise to a large giant component, whereas teams publishing in low-impact journals typically form small isolated clusters. This suggests that teams publishing in high-impact journals perform a better sampling of the knowledge within a field and thus are able to more efficiently use the resources of the invisible college. Surprisingly, neither p , q , or S were significantly correlated with impact factor in astronomy. This distinguishes astronomy from the other creative enterprises considered.

We have shown that team size evolves with time, probably up to an optimal size as in the case of the BMI. A similar process may be occurring for the parameters quantifying expertise, p , and diversity, q . Four of the five fields considered, all except astronomy, have very similar values of p and q , thus suggesting that a “universal” set of

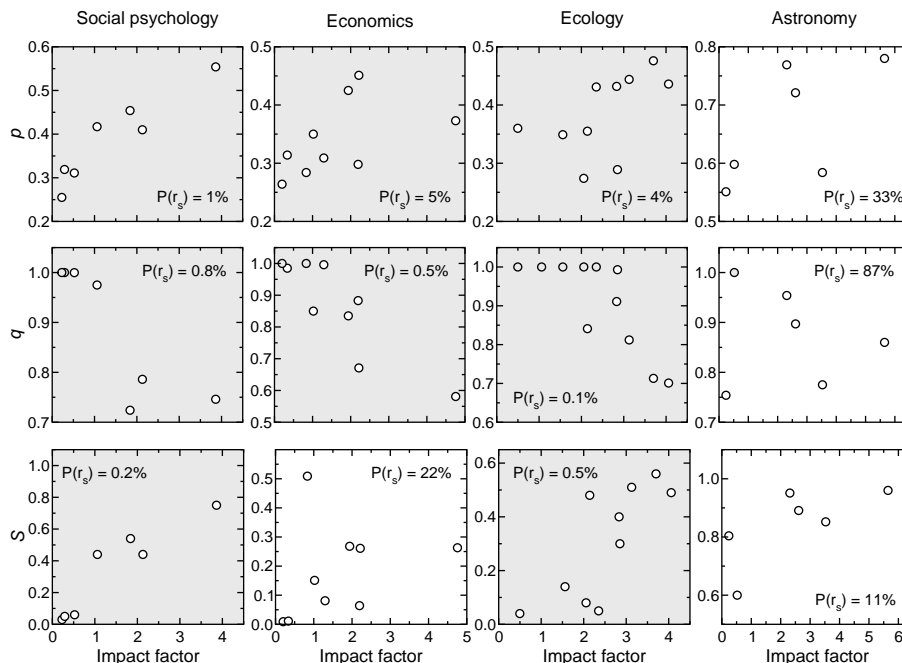


Fig. 5. Relation between team assembly mechanisms, network structure, and performance. We calculate the values of p , q , and S for several journals in each of the four scientific fields considered. In a few cases, q should be larger than one in order to reproduce the empirical values of f_R ; in these cases, q is considered one and the corresponding points are shaded. We plot the values of p , q , and S as a function of the impact factor of the journal and then use the Spearman rank-order correlation coefficient r_s to determine significant correlations. Shaded graphs indicate significantly correlated variables at the 95% confidence level.

optimal values might exist. The fact that in astronomy there are no correlations between p , q , or S and the impact of journals also indicates that this field is different from the others. Whether these differences are caused by the needs imposed by the creative enterprise itself or to historical or other reasons is a question that we cannot answer conclusively.

References and Notes

1. M. S. Granovetter, *Am. J. Sociol.* **78**, 1360 (1973).
2. R. Reagans, E. W. Zuckerman, *Organ. Sci.* **12**, 502 (2001).
3. R. Burt, *Am. J. Sociol.* **110**, 349 (2004).
4. B. Uzzi, J. Spiro, *Am. J. Sociol.*, in press.
5. J. R. Larson, C. Christensen, A. S. Abbott, T. M. Franz, *J. Pers. Soc. Psychol.* **71**, 315 (1996).
6. A. Edmondson, *Adm. Sci. Q.* **44**, 350 (1999).
7. K. A. Jehn, G. B. Northcraft, M. A. Neale, *Adm. Sci. Q.* **44**, 741 (1999).
8. G. Stasser, D. D. Stewart, G. M. Wittenbaum, *J. Exp. Soc. Psychol.* **31**, 244 (1995).
9. J. R. Katzenback, D. K. Smith, *The Wisdom of Teams* (Harper Business, New York, 1993).
10. J. M. Ziman, *Prometheus Bound* (Cambridge Univ. Press, Cambridge, 1994).
11. J. R. Brown, *Science* **290**, 1701 (2000).
12. S. Green, K. Green, *Broadway Musicals Show by Show* (Hal Leonard, Milwaukee, WI, 1996), ed. 5.
13. R. Simas, *The Musicals No One Came to See: A Guidebook to Four Decades of Musical-Comedy Casualties on Broadway, Off-Broadway and in Out-Of-Town Try-Out, 1943-1983* (Garland, New York, 1988).
14. We imposed several requirements on the journals we selected for analysis. First, the main subject category of the journal must be the desired one. For example, we consider only those ecology journals whose subject category is either ecology or ecology and biodiversity and conservation according to the journal

- Citation Reports. We disregarded more specialized journals, such as *Microbial Biology*, whose subject category is more specific. We also required that journals contain a sufficiently large number of papers, typically larger than 1000.
15. M. E. J. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 404 (2001).
16. A.-L. Barabási *et al.*, *Physica A* **311**, 590 (2002).
17. M. E. J. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5200 (2004).
18. K. Börner, J. T. Maru, R. L. Goldstone, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5266 (2004).
19. J. J. Ramasco, S. N. Dorogovtsev, R. Pastor-Satorras, *Phys. Rev. E* **70**, 036106 (2004).
20. This stationary state remains until the mid-1980s, when size drops again precisely at the time when a rash of revivals and revues conceivably simplified production.
21. A.-L. Barabasi, R. Albert, *Science* **286**, 509 (1999).
22. D. J. Watts, S. H. Strogatz, *Nature* **393**, 440 (1998).
23. L. A. N. Amaral, A. Scala, M. Barthélémy, H. E. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11149 (2000).
24. R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
25. M. E. J. Newman, *SIAM Rev.* **45**, 167 (2003).
26. L. A. N. Amaral, J. Ottino, *Eur. Phys. J. B* **38**, 147 (2004).
27. H. Etkowitz, C. Kemelgor, M. Neuschatz, B. Uzzi, J. Alonzo, *Science* **266**, 51 (1994).
28. D. A. Harrison, K. H. Price, M. P. Bell, *Acad. Manage. J.* **41**, 96 (1998).
29. S. G. Barsade, A. J. Ward, J. D. F. Turner, J. A. Sonnenfeld, *Adm. Sci. Q.* **46**, 174 (2001).
30. The teams and the agents are the nodes in a bipartite network. Technically, agents are connected only to teams and vice versa. However, this bipartite network can be projected onto a network comprising only agents and in which there is an edge (connection) between two nodes (agents) if the agents have been connected to at least one common team.
31. D. J. de Solla Price, *Little Science, Big Science... and Beyond* (Columbia Univ. Press, New York, 1963).

32. R. K. Merton, *The Sociology of Science* (Univ. of Chicago Press, Chicago, 1973).
33. D. Stauffer, A. Aharony, *Introduction to Percolation Theory* (Taylor and Francis, ed. 2, London, 1992).
34. M. Gladwell, *The Tipping Point: How Little Things Can Make a Big Difference* (Little, Brown, Boston, 2000).
35. Figure 3C shows that large f_R occurs when p and q are large and corresponds to a network in which collaborations among incumbents are firmly established and opportunities for newcomers are few. Conversely, small f_R , which occurs when p and/or q are small, indicates plentiful opportunities for newcomers to join new projects. In this case, newcomers are the norm and collaborations are rarely repeated. Lastly, intermediate values of f_R suggest intermediate values of both p and q , that is, a situation for which there is a balance between seasoned incumbents and newcomers with fresh ideas.
36. The value of p is directly given by the fraction of incumbents in new creations. The value of q must be obtained numerically by simulating the model with different tentative values of q until the fraction f_R of repeat incumbent-incumbent links predicted by the model coincides with the value measured from the data.
37. We thank K. Börner, V. Hatzimanikatis, A. A. Moreira, J. M. Ottino, M. Sales-Pardo, and D. B. Stouffer for numerous suggestions and discussions. R.G. thanks the Fulbright Program and the Spanish Ministry of Education, Culture, and Sports. L.A.N.A. gratefully acknowledges the support of a Searle Leadership Fund Award and of a NIH/National Institute of General Medical Studies K-25 award.

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5722/697/DC1
Materials and Methods
Figs. S1 and S2

13 October 2004; accepted 10 February 2005
10.1126/science.1106340

The Dynamics of Interhemispheric Compensatory Processes in Mental Imagery

A. T. Sack,^{1*} J. A. Camprodon,² A. Pascual-Leone,² R. Goebel¹

The capacity to generate and analyze mental visual images is essential for many cognitive abilities. We combined triple-pulse transcranial magnetic stimulation (tpTMS) and repetitive TMS (rTMS) to determine which distinct aspect of mental imagery is carried out by the left and right parietal lobe and to reveal interhemispheric compensatory interactions. The left parietal lobe was predominant in generating mental images, whereas the right parietal lobe was specialized in the spatial comparison of the imagined content. Furthermore, in case of an rTMS-induced left parietal lesion, the right parietal cortex could immediately compensate such a left parietal disruption by taking over the specific function of the left hemisphere.

Mental imagery refers to the experience of a perception in the absence of a corresponding physical stimulus. In our everyday life, mental imagery represents a crucial element of numerous cognitive abilities, such as object recognition, reasoning, language comprehension, and memory. Because of its importance, the exact processes associated with imagery have long occupied cognitive psychologists and been a matter of debate and controversy (1).

Mental imagery is accompanied by the activation of frontoparietal networks (2–5), but the exact brain areas engaged in imagery depend on the specific features of the imagery task (6). When spatial comparisons between imagined objects are required, most functional imaging studies show bilateral parietal activation in homologous intraparietal sulcus areas of the left and right hemispheres (3). However, neuropsychological studies on patients with focal brain lesions generally support a domi-

nant role of the left hemisphere in imagery [(7), but see (8)].

Time-resolved functional magnetic resonance imaging (fMRI) has been used to address this apparent contradiction between functional imaging studies and findings in focal brain injury patients (4). An earlier cluster of activation in both parietal cortices (with left predominance) can be separated from a late cluster confined to the right parietal cortex (Fig. 1). These results support the involvement of both parietal lobes in mental imagery but suggest that each parietal lobe has a distinct functional role at different moments in time. The sequential parietal activation might represent a transition from an earlier more distributed processing stage of image generation to a later right-hemispheric lateralized stage of spatial analysis of the images (4). In a combined fMRI and rTMS study, only rTMS to the right parietal lobe led to an impairment of spatial imagery performance

¹Department of Cognitive Neuroscience, Faculty of Psychology, Maastricht University, Post Office Box 616, 6200 MD Maastricht, Netherlands. ²Center for Non-invasive Brain Stimulation, Beth Israel Deaconess Medical Center and Harvard Medical School, 330 Brookline Avenue, Kirstein Building KS 452, Boston, MA 02215, USA.

*To whom correspondence should be addressed. E-mail: a.sack@psychology.unimaas.nl

during and immediately after rTMS (5). Hence, only the right parietal activity seems to be functionally relevant for spatial imagery, whereas the rTMS-induced disruption of the left parietal activity has no measurable impact on the ability to perform the imagery task.

However, the behavioral consequences of focal disruption of brain activity (by injury or transiently by rTMS) show the ability of the rest of the brain to cope with the insult. Therefore, the apparently unaffected imagery performance after left parietal disruption could be due to compensatory processes within a distributed neural network that cannot be sustained after right parietal disruption. We hypothesized that the right parietal cortex is critical for the spatial comparison of mental images but is also able to compensate a disruption of left parietal activity during image generation. This assumes that in case of a left parietal lesion, the right parietal cortex will subserve both functions, image generation and spatial comparison.

We introduced a previously untested TMS procedure, which combines the respective advantages of creating a temporary virtual lesion by rTMS with the precise chronometric study offered by event-related tpTMS (9). We used tpTMS to chart the time points at which right parietal activity is critical for spatial imagery, depending on whether the homologous left parietal cortex had or had not been suppressed by preceding rTMS (Fig. 2).

Subjects were asked to imagine two analog clock faces based on acoustically presented times (e.g., two o'clock and five o'clock) and to judge at which of the two times the clock hands form the greater angle (4, 5, 10). This task requires the subjects to first generate visual mental images of two analog clocks and then mentally compare the angles formed by the clock hands (11). The different mental processes associated with the mental clock task include the encoding of the acoustic stimuli, the generation and maintenance of the mental images, the spatial comparison, and the decision and response by button press (Fig. 1). We hypothesized that tpTMS over the right parietal cortex would only impair performance when applied at a specific late time point during the task (representing the spatial comparison stage). Moreover, if the right parietal cortex can compensate a disruption of left parietal activity (representing image generation), a functional lesion of the left parietal cortex by means of preceding rTMS should result in an extension of this late critical time point of right parietal activity toward an earlier time point. This implies that in case of a left parietal lesion, the right parietal cortex is able to subserve image generation and spatial comparison (Fig. 2).

Every subject completed four right parietal tpTMS runs, two with and two without pre-

ceding left parietal rTMS (rTMS factor) (12). A within-subject repeated measures analysis of variance with the two-level rTMS factor and a four-level time factor of right parietal disruption revealed a significant interaction ($F = 6.358$; $P = 0.05$). More specifically, in the runs without preceding rTMS over the left parietal cortex (intact left parietal lobe), tpTMS over the right parietal cortex only significantly impaired the behavioral performance in the mental clock task when applied at a time interval of 4200 ms after trial onset (13), which fell within time bin 3 (spatial comparison; Student's t test = 4.099, $df = 5$, $P < 0.05$; Fig. 3; individual data available in fig. S1). This highly specific and rather late time point corresponds well to the highly lateralized right

parietal activity cluster previously revealed by time-resolved fMRI (4). In contrast, in the runs with preceding rTMS (i.e., during the temporary disruption of left parietal cortical function), the temporal characteristics of the behavioral effects of tpTMS to right parietal cortex changed significantly and now not only included time bin 3 (spatial comparison; $t = 4.511$, $df = 5$, $P < 0.05$) but also several earlier time intervals (14) during time bin 2 (image generation; $t = 6.217$, $df = 5$, $P < 0.05$; Fig. 3; see also fig. S1 for individual data). These earlier time windows correspond well to the bilaterally distributed parietal activation cluster revealed by time-resolved fMRI (4). Analysis of the error rates revealed no significant differences between the different time

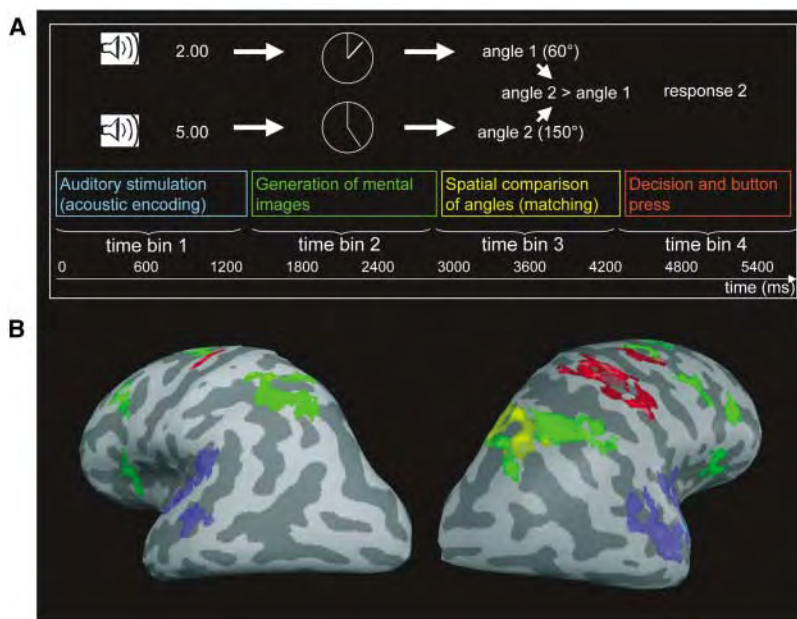


Fig. 1. Brain areas activated during spatial imagery. Temporal sequence of cognitive processes associated with the mental clock task (A) and the activated brain areas as revealed by time-resolved fMRI based on a different subject sample from a previous study. (B) View from the occipital pole of both inflated hemispheres. The color code from blue, to green, to yellow, to red indicates the sequence of activated brain areas associated with the four required cognitive processes. The results distinguish an earlier, bilateral parietal activation (green) from a later, strictly right parietal activation (yellow).

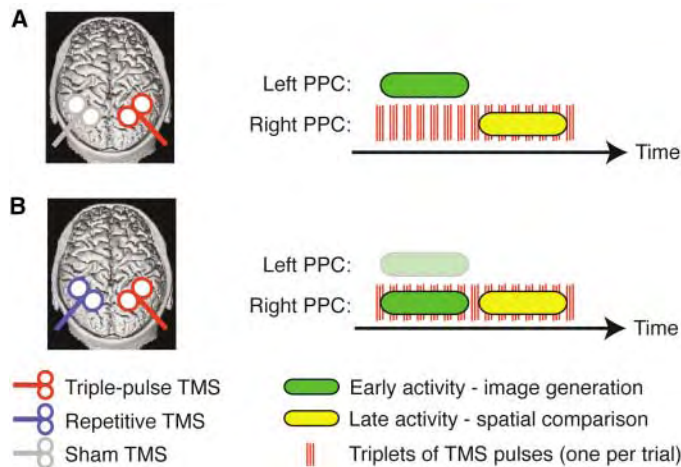


Fig. 2. TMS protocol. tpTMS (red coil) identifies the time points at which right parietal activity is critical for spatial imagery (A) after sham TMS has been applied to the left parietal cortex (gray coil) and (B) after inducing a virtual lesion to the left parietal cortex by preceding rTMS (blue coil). PPC, posterior parietal cortex.

⊗ Triple-pulse TMS ▭ Early activity - image generation
⊗ Repetitive TMS ▭ Late activity - spatial comparison
⊗ Sham TMS ||| Triplets of TMS pulses (one per trial)

points and/or conditions (fig. S2). In an additional control experiment, left parietal rTMS alone, without subsequent right parietal tpTMS, led to no significant impairments in the task performance of the mental clock task ($t = 0.601$; $df = 5$; $P = 0.574$).

Based on the temporal activation sequence from left to right parietal cortex during an imagery task that involves the generation as well as spatial comparison of mental images (4), we applied tpTMS to fractionate specialized processing components in the right parietal cortex. We were able to show that an rTMS-induced unilateral disruption can lead to a significant change in the critical functional time point, at which activations in the homologous cortical area of the contralateral hemisphere are crucial for the execution of a specific function. Whereas the mental process of spatial comparison is highly lateralized to the right hemisphere, image generation shows only a weak left-hemispheric lateralization with a more bilateral distribution. By charting and comparing the time courses of critical right parietal activity with and without preceding left parietal disruption, our data suggest that the right hemisphere is able to compensate for (virtual) lesions of the left hemisphere by taking over

this specific mental process. Whereas an rTMS-induced left parietal disruption alone did not impair task performance, a subsequent right parietal tpTMS early in the task (during image generation) unmasked the behavioral deficit, most likely by blocking the right parietal compensation.

Discrepancies across studies concerning the hemispheric lateralization during mental imagery likely arise, because different aspects of imagery are carried out by different parts of a bihemispheric neural network. Our results are in accordance with neuropsychological models of spatial imagery, which propose that the generation of mental images relies primarily on structures in the posterior left hemisphere, whereas spatial operations on these images are subserved by the posterior right hemisphere (15). Moreover, our results reveal that in case of left parietal disruption, the generation of mental images might be carried out by the right hemisphere as part of an asymmetric inter-hemispheric compensatory mechanism. An isolated deficit of the ability to generate inner visual images after unilateral lesion is clinically hardly ever reported (8), which could be explained on the basis of the compensatory processes revealed in our study. This also parallels evidence from spatial hemineglect, which mostly occurs after right hemispheric lesions. This phenomenon has been explained on the basis of an asymmetrical distribution of spatial attention (16), in which a right parietal lesion would lead to hemineglect for the left visual field, whereas a respective left parietal lesion could be compensated for by the right hemisphere. However, also on the basis of such an asymmetrical distribution of inter-hemispheric attention, it cannot be assumed that the two hemispheres simply process information independently. Human cognition rather includes highly complex processes of interhemispheric competition, cooperation, and suppression. It remains speculative whether the revealed functional compensation represents a real cortical reorganization or whether the right hemisphere compensates the behavioral consequence due to the bilateral distribution of image generation; our results are in line with the latter. In this respect, the mental process of image generation might indeed be carried out by both hemispheres. However, under normal physiological circumstances, the functional necessity of the right hemisphere becomes redundant because of the left hemispheric dominance. Hence, the revealed compensatory processes might be based on a disruption of the suppressive influence of one hemisphere within an inter-hemispheric competition. The release of this suppressive influence by rTMS would then restore the functional relevance of the contralateral hemisphere, thereby compensating or even enhancing the cognitive functions it subserves (17).

References and Notes

1. S. M. Kosslyn, *Image and Mind* (Harvard Univ. Press, Cambridge, MA, 1980).
2. E. Mellet et al., *J. Neurosci.* **16**, 6504 (1996).
3. L. Trojano et al., *Cereb. Cortex* **10**, 473 (2000).
4. E. Formisano et al., *Neuron* **35**, 185 (2002).
5. A. T. Sack et al., *Neuron* **35**, 195 (2002).
6. S. M. Kosslyn, M. Behrmann, M. Jeannerod, *Neuropsychologia* **33**, 1335 (1995).
7. M. J. Farah, M. S. Gazzaniga, J. Holtzman, S. M. Kosslyn, *Neuropsychologia* **23**, 115 (1985).
8. L. Trojano, D. Grossi, *Brain Cognit.* **24**, 213 (1994).
9. In event-related TMS, a magnetic pulse is delivered at a precise time point during task execution, leading to a disruption of the targeted brain region for less than 100 ms (18, 19). By applying single-pulse TMS at variable times during task execution, it is possible to achieve a time resolution of 5 to 10 ms (20, 27). In the context of a multicomponential task, the tpTMS protocol introduced in this study allows us to identify the time point at which the targeted neuronal activity is functionally critical for the execution of a particular mental subprocess. We assume that applying tpTMS at a functionally critical time point leads to a disruption of the respective mental subprocess executed at that time, resulting in the necessity to restart the mental processes after the effect of the TMS triple has subsided. In rTMS, a train of pulses is delivered at a given stimulation frequency to a given brain region, modulating the excitability of the stimulated area beyond the duration of the rTMS application itself (20, 22).
10. A. Paivio, *J. Exp. Psychol. Hum. Percept. Perform.* **4**, 61 (1978).
11. In every trial, two clock times were acoustically presented in rapid succession. Each trial lasted for 2000 ms (1000 ms per clock time). Subjects had to press the left mouse button if the hands of the first imagined clock formed the greater angle or the right mouse button for the second. Subjects were instructed to respond as accurately as possible. Stimuli involved only half-hour or hour intervals and were balanced for the spatial side of the clock on which the hands had to be imagined as well as the numerical greatness of the corresponding digital time.
12. Materials and methods are available as supporting material on Science Online.
13. Mean percentage change = 46.6%; $t = 3.693$; $df = 5$; P (uncorrected) = 0.014.
14. Time windows were at 1500 ms [mean percentage change = 31.5%; $t = 4.836$; $df = 5$; P (uncorrected) = 0.005] and 2100 ms [mean percentage change = 30.4%; $t = 4.282$; $df = 5$; P (uncorrected) = 0.008] after trial onset, including also one early time window from 600 ms [mean percentage change = 31.5%; $t = 4.637$; $df = 5$; P (uncorrected) = 0.006] to 900 ms [mean percentage change = 25.0%; $t = 2.903$; $df = 5$; P (uncorrected) = 0.034].
15. M. J. Farah, *Trends Neurosci.* **12**, 395 (1989).
16. M. M. Mesulam, *Philos. Trans. R. Soc. London Ser. B* **354**, 1325 (1999).
17. C. C. Hilgetag, H. Theoret, A. Pascual-Leone, *Nat. Neurosci.* **4**, 953 (2001).
18. V. Moliadze, Y. Zhao, U. Eysel, K. Funke, *J. Physiol.* **553**, 665 (2003).
19. G. Thut et al., *Clin. Neurophysiol.* **114**, 2071 (2003).
20. E. M. Robertson, H. Theoret, A. Pascual-Leone, *J. Cognit. Neurosci.* **15**, 948 (2003).
21. V. Walsh, A. Cowey, *Nat. Rev.* **1**, 73 (2000).
22. V. Walsh, A. Pascual-Leone, *Neurochronometrics of Mind: Transcranial Magnetic Stimulation in Cognitive Science* (MIT Press, Cambridge, MA, 2003).
23. Funded by the Netherlands Organization for Scientific Research (NWO 451-03-038) and the NIH (K24 RR018875, NCR01 MO1 RR01032, and RO1MH60734). We thank P. de Weerd, E. Formisano, and A. Kaas for helpful comments on the manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/308/5722/702/DC1

Materials and Methods
Figs. S1 and S2

22 November 2004; accepted 20 January 2005
10.1126/science.1107784

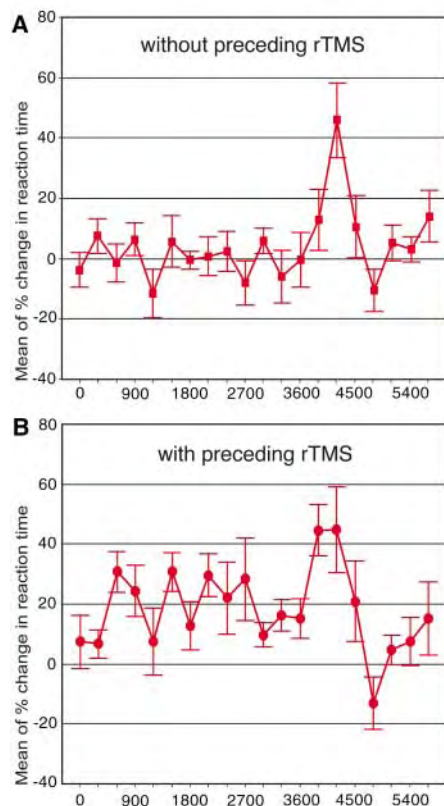


Fig. 3. Behavioral results. Mean percentage change of reaction time during tpTMS over the right parietal cortex at the different time intervals, independently for the two rTMS conditions: (A) without preceding left parietal rTMS and (B) with preceding left parietal rTMS.

NEW PRODUCTS

<http://science.labvelocity.com>

Genotyping System

The GenomeLab SNPstream Genotyping System is a scalable solution to meet medium- to high-throughput needs. This automated, multiplexed system can process up to 48 single nucleotide polymorphisms (SNPs) in each well of an arrayed 384-well plate. SNPstream operates at a consistent cost per genotype or per sample, regardless of run throughput, allowing users to cost-effectively zero in on targets with low-throughput studies as well as conduct high-throughput analyses. Based on single-based primer extension technology, the system operates at 99% accuracy. The system delivers accurate results using as little as 2 ng of genomic DNA. The plates are compatible with Biomek series liquid handlers, for fast pre- and post-polymerase chain reaction setup.

Beckman Coulter For information 800-742-2345 www.beckmancoulter.com

Protein Purification Columns

New HisTap FF crude prepacked columns (1 ml and 5 ml) are designed to meet the growing need for pure proteins to use in functional and structural genomics and proteomics research. These new columns enable fast, reliable, and simple one-step purification of histidine-tagged recombinant proteins directly from unclarified cell lysates. The samples minimize the need for sample pre-treatment. Sonicated cell lysate can be directly loaded onto the column, eliminating the need for a clarification step (centrifugation or filtration).

GE Healthcare For information +44 (0) 1494 498 068 www.gehealthcare.com

Monoclonal Antibodies

Covance's Custom Immunology Services provide high-quality monoclonal antibodies to the life science industry and research community. Covance can meet the challenge of the most difficult hybridoma development projects for the production of monoclonal antibodies. Covance offers cell line optimization, recloning, mycoplasma testing and decontamination, downstream bioprocess development, and cell culture production in addition to hybridoma development services.

Covance For information 800-345-4114 www.CRPinc.com

Non-Treated Tissue Culture

Non-Treated Tissue Culture Multidishes and EasYFlasks are designed for research on living cells in which attachment is not needed. Applications include the growth of hybridoma cell lines and suspension cell cultures, or when the hydrophilic and cell-binding characteristics of the traditional Nunclon Δ surface are too strong. The non-treated products are manufactured without the active Nunclon Δ surface that enhances cell attachment. Non-treated EasYFlasks are designed to permit maximum airflow under each flask for optimal temperature equilibration during incubations, even when stacked.

Nunc For information 866-686-2548 www.nuncbrand.com

Heat Shock Proteins

Almost four dozen stress proteins and related antibodies are available. Present in all cells, these proteins function as molecular chaperones that assist in the folding and unfolding of macro-

molecules and in the assembly and disassembly of higher order structures. They have also been shown to play a role in resistance to cellular stresses such as heat, cold, heavy metals, oxygen deprivation, and ischemia. Stress proteins may also be involved in the immune response and hold promise as a potential cancer therapy.

Axxora For information 800-550-3033 www.axxora.com

Two-Dimensional Gel Scanner

The ProteomeScan high-resolution image scanner can generate high-quality, two-dimensional (2D) gel images in seconds for scientists looking for accuracy in proteomics research. The ProteomeScan is controlled by icon-driven software and allows users to dictate the kind of image they would like by selecting scanning resolution, compression, and scan speed. The system can generate images of up to 12,800 \times 12,800-dpi resolution, ensuring that even the smallest protein spots are detected. For scientists wanting to save time with their 2D gel analysis, the ProteomeScan can transfer images directly into Dymension, Syngene's new 2D image analysis software, which can analyze a typical 2D gel image in seconds. The ProteomeScan can also switch between transmittance and reflectance modes and has variable color scanning, which means it is easy to scan a variety of different gel stains, such as silver, Coomassie blue, and Sypro Ruby. There are two versions available, for gels up to 21 cm by 29 cm and 31 cm by 43 cm.

Syngene For information 800-686-4407 www.syngene.com



Fluorescence Membranes

The Immobilon-FL membrane is a protein immunoblotting membrane optimized for fluorescence applications. The background fluorescence of this polyvinylidene fluoride (PVDF) microporous membrane is nearly 10 times lower than typical PVDF membranes and two to five times lower than nitrocellulose membranes. As a result, the new membrane offers greater sensitivity and improved protein detection with fluorescent reagents. The Immobilon-FL membrane has a 0.45 μ m pore size, which is recommended for blotting proteins greater than 20 kDa. It can be used at all excitation and emission wavelengths in both the visible and infrared ranges, and can be used with chromogenic or chemiluminescent detection methods. It is part of a family of PVDF blotting membranes, including Immobilon-P for general applications and Immobilon-PSQ for blotting small (less than 20 kDa) proteins.

Millipore For information 800-MILLIPORE www.millipore.com/immobilon

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and government organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier by visiting www.science.labvelocity.com on the Web, where you can request that the information be sent to you by e-mail, fax, mail, or telephone.

For more information visit **GetInfo**,
Science's new online product index at
<http://science.labvelocity.com>

From the pages of GetInfo, you can:

- Quickly find and request free information on products and services found in the pages of *Science*.
- Ask vendors to contact you with more information.
- Link directly to vendors' Web sites.

Not Your Father's Postdoc

In today's scientific labor market, just doing good science is no longer enough. Postdocs need realistic expectations, good information, and an entrepreneurial attitude toward their careers

On 18 March, 100 years and a day after 26-year-old Albert Einstein sent off the first of his 1905 papers that were destined to revolutionize physics, National Institutes of Health (NIH) Director Elias Zerhouni invoked the name of another Nobelist, biochemist Marshall Nirenberg, at a meeting held to unveil a new report on the plight of young researchers today. Nirenberg won his Nobel Prize at 41—even younger than Einstein. “In today’s world,” Zerhouni noted, “Marshall Nirenberg would get his Nobel Prize before he got his first NIH grant.”

Today’s young biomedical researchers, notes the National Research Council (NRC) report *Bridges to Independence*, don’t win their first independent faculty appointment until a median age of 36, and they don’t reach the milestone that marks their *real* debut as independent investigators—their first competitive NIH research grant—until a median age of 42. This late start doesn’t just stunt individual careers, warns the report. It also threatens the vitality of the nation’s scientific enterprise.

Moreover, for most aspiring biomedical scientists, there won’t be an academic job at the end of that long apprenticeship: There are simply far more people in the pipeline than there are available academic positions. Given that inescapable arithmetic, experts advise today’s budding biomedical Einsteins and Nirenbergs to think more broadly about their future scientific careers. “The number-one thing that every postdoc needs to think about is what they want to do when they grow up,” says Ida Chow, executive officer of the Society of Developmental Biology.

Funding patterns and holding patterns

NIH funding is itself largely responsible for the slowdown, explains the report. Over the past several decades, NIH has financed a swift rise in the number of life science Ph.D.s and then supported them—mostly by means of extramural research grants made to universities—in postdoctoral appointments that have become, in the report’s words, “a ‘holding pattern’ for thousands of young scientists” who find themselves unable to move on to traditional faculty

of biomedical researchers.” Between 1993 and 2000, the number of U.S. life science Ph.D.s under age 35 holding coveted tenure-track jobs in major research universities declined by 12.1%, to 543; meanwhile, the number of U.S. biomedical Ph.D.s in that age range increased by 59%, to nearly 20,000, and tens of thousands more scientists with foreign Ph.D.s came to fill postdoc positions in U.S. labs.

The traditional “linear progression” from “graduate school to postdoctoral positions to assistant professorships, then obtaining funding and tenure” now works for only a small minority of young scientists, the report explains. Instead of this simple progression,

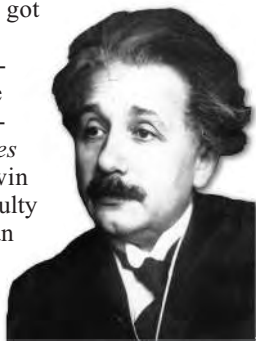
young scientists confront “a complex network of current career pathways” to a variety of occupations using scientific training, many of them outside academe. In addition, increasing numbers of scientists hold non-tenure-track university posts, a type of appointment that increased 55% between 1990 and 2001, a rate approximately seven times faster than that of tenure-track posts. The great majority of postdocs seeking stable career employment must therefore take what the academic world has long regarded as “alternative” jobs with unfamiliar professional cultures and skill requirements that scientists generally do not encounter in graduate school or a mentor’s lab.

Building bridges to opportunity

In light of these changes, what can post-

docs do to prepare themselves to move beyond the training phase, wherever that move might take them? A first step is to jettison the notion of jobs outside academe as “alternative” work, advises Chow. “The word ‘alternative’ gives a bad connotation of second class,” she says. A far better term, she believes, is “career choices,” specifically “the many career choices that science graduates—from the bachelors to the doctorate—have today compared to a generation ago.”

Opportunities include industry—which in 2001 employed some 35% of life science



Albert Einstein
b. 1879

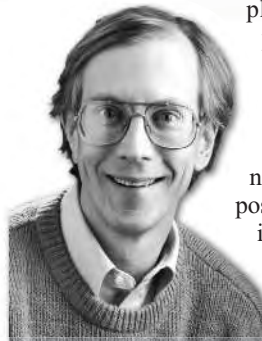
Age	Event
23	Began work at Patent Office
26	<i>Annus mirabilis</i>
32	1st permanent post
42	Won Nobel Prize

Researchers today win their first competitive NIH grant at a median age of 42. These three scientists won Nobel Prizes at that age.



Marshall Nirenberg
b. 1927

Age	Event
21	Received B.S. in zoology from University of Florida, Gainesville
25	M.Sc. in zoology from University of Florida
30	Ph.D. in biochemistry from University of Michigan, Ann Arbor
32	Postdoc at NIH
33	Appointed research biochemist at NIH
35	Made section head, Biomedical Genetics, NIH
38	Won National Medal of Science
41	Won Nobel Prize



Thomas Cech
b. 1947

Age	Event
19	Entered Grinnell College
23	Started grad school at UC Berkeley
28	Earned Ph.D. from UC Berkeley
	Started postdoc at MIT
31	1st faculty position (University of Colorado, Boulder)
41	Appointed Howard Hughes Medical Institute Investigator
	Albert Lasker Award in Basic Medical Research
42	1989 Nobel Prize in chemistry

posts. The postdoc—a de facto requirement for an academic research career—now averages just under 5 years. For many life science postdocs, especially among the 80% paid out of NIH grants to principal investigators, “‘postdoctoral training’... has turned into ‘postdoctoral employment’—with the postdoc remaining at the same professional position with little advancement of professional training,” the report says.

“Simply put,” notes the report in a model of understatement, “there are not enough tenure-track academic positions for the available pool

Ph.D.s, up from 15% in 1981—as well as government, science policy, writing, and nonuniversity teaching. “Even Wall Street needs people with science backgrounds to work as analysts,” Chow says. “There are many more choices than just university jobs.”

Setting a personal course for the future is particularly important at the postdoc stage, when young scientists no longer have the structure and goals automatically supplied by graduate school, says Andrea Stith, science policy analyst at the Federation of American Societies for Experimental Biology (FASEB): “There your goal is defined for you, and you have the evaluation of grades.”

At one time, most postdocs’ goals were also clear—a faculty job—and the guidance and help of the PI played a major role in getting there. But as the range of jobs scientists occupy has expanded, the help their advisers can provide has diminished. Faculty members who have spent their careers within academe often lack the knowledge and contacts needed to help their protégés find jobs in other sectors. So postdocs considering opportunities outside academic science need to assume far greater responsibility for their own futures. Key to taking charge, says Stith, is systematic planning. An effective approach to doing so, Stith continues, is to create an Individual Development Plan (IDP), a document that states specific goals and outlines specific means of achieving them.

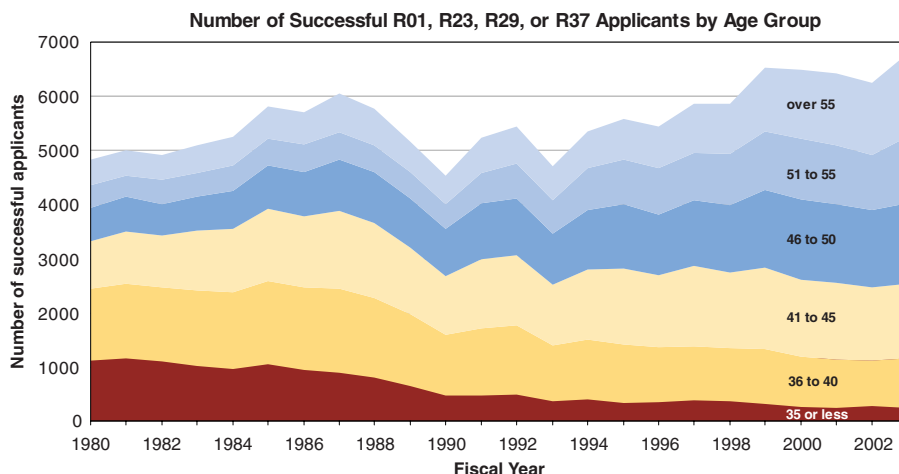
Doing science on yourself

Widely used in the business world, the IDP is unfamiliar to most academic scientists, although some universities and funders now use IDPs to help plan the postdoctoral period. FASEB has developed a 3-hour instructional IDP seminar, complete with interactive exercises, that it piloted at the Experimental Biology 2005 meeting in April in San Diego, California, and plans to present at other venues. “An IDP is appropriate for every stage of your life,” says Stith, who served as one of the seminar presenters.

Like doing an experiment, the four-step process of creating an IDP involves thinking strategically, gathering data, and evaluating results. It begins with a self-assessment during which the individual determines his or her own values, interests, preferences,

priorities, strengths, weaknesses, talents, and tolerances. “Is the amount of pay important to you?” Stith asks. “Is time with family? Is independence, as far as determining the project you’re working on? How much and what are you going to compromise?”

Next comes the career-assessment stage, when the postdoc identifies and learns about occupations that appear to meet his or her needs. Information gathered should include the skills, knowledge, and characteristics needed to enter and succeed in the fields of interest and how to go about acquiring them. Sources of information can include university career centers and postdoc offices, professional associations, libraries, the Internet, and networking with people who have firsthand experience.



Delayed independence. Researchers under 40 now account for less than 15% of NIH grant awards.

In the third stage, the postdoc composes the document. “Write down your goals and parse out your long-term and short-term goals,” Stith says. Explicit timelines add specificity. Finally, in stage four, the individual puts the plan into effect, periodically measuring progress toward those goals and revising the plan as needed. “We expect people’s interests to change,” Stith says.

Weaning or weeding?

The entrepreneurial spirit symbolized by FASEB’s IDP could be particularly handy for postdocs in the next few years if the recommendations in *Bridges to Independence* are adopted. The recommendations would create opportunities for some postdocs and insecurity for others, allowing—indeed, forcing—many postdocs to “grow up” to some form of independence more quickly.

On the opportunities side, one prominent proposal would reallocate NIH research funds to hundreds of new awards each year to postdocs doing their own research. Another would strengthen support for the growing cadre of investigators on soft money in non-tenure-track positions. Yet even the downside of these

proposals is likely to have a secondary weaning effect: Given current budget constraints, these initiatives would most likely take funding away from some current investigators—and paychecks from their postdocs.

From a postdoc point of view, perhaps the most significant recommendation is one that would limit to a total of 5 years the postdoctoral support any individual could receive from all NIH sources combined, whether fellowships or employment on PI grants. This would eject the longest-serving postdocs from their current jobs and could endanger the immigration status of many noncitizens, who account for more than half of the postdocs working in U.S. labs. *Bridges* urges PIs to promote scientists remaining on after their NIH eligibility ends to staff

positions with pay, benefits, and clearly defined institutional status commensurate with their experience and responsibilities. But doing so would be expensive, and the source of money to support it is unclear.

The weeding out that would occur is consistent with the goals of the NRC committee. “This is not a full employment system for postdocs,” said National Academy of Sciences president Bruce Alberts at the 18 March briefing. “The system will select out those of real ability [so that] the very best have a chance to see what they can do.” The changes “might be painful to some people,” acknowledged the report committee’s chair, Thomas Cech, president of the Howard Hughes Medical Institute in Chevy Chase, Maryland—who, incidentally, won the Nobel Prize in chemistry just as he turned 42—but they “should have a wonderful effect on encouraging early consideration of career opportunities.”

Whether or not these recommendations are adopted—and the report itself points out that earlier recommendations were not—the career picture for most postdocs remains complex for the foreseeable future. “Each year, both new and experienced investigators compete in a Darwinian-like system,” the *Bridges* report states. It should therefore come as no surprise to life scientists that those who adapt strategically to rapidly changing circumstances have the best chance of prospering in the years ahead.

—BERYL LIEFF BENDERLY

Beryl Lieff Benderly is a contributor to *Science’s* Next Wave (www.nextwave.org).

SOURCE: NIH